

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:44:15 ; Search time 34.4229 Seconds  
(without alignments)  
225.942 Million cell updates/sec

Title: US-09-423-100-1

Perfect score: 260

Sequence: 1 MFPTIPLSLKFDNAMLRAHR.....QEFEEYIPKQKYSFLQNP 49

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 11364

Minimum DB seq length: 3  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

1: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
2: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
3: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
4: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
5: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
6: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
7: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
8: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
9: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
10: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
11: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
12: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
13: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
14: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
15: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
16: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
17: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
18: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
19: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
20: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*  
21: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*  
23: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA2002.DAT.\*  
24: /SIDSL/qcgdata/geneseq/geneseqp-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length DB	ID	Description
1	260	100.0	49	AA142855	Human growth hormo
2	260	100.0	92	AA142856	Human growth hormo
3	260	100.0	107	AA142860	hGH-mini-proinsulin
4	260	100.0	134	AA192265	Human anti-angioge
5	260	100.0	140	AA191041	Human growth hormo
6	260	100.0	150	AA142861	Chimeric protein,
7	260	100.0	192	AA190129	Human growth hormo
8	260	100.0	192	AA192264	Human anti-angioge
9	260	100.0	261	AA191299	Human nerve growth

10	260	100.0	262	7	AA161033	Human beta-nerve q
11	260	100.0	262	12	AA11740	Human growth hormo
12	260	100.0	310	11	AA103255	Fusion protein of
13	257	96.8	344	11	AA105313	Segment of B-cell
14	256	98.5	204	23	AA177327	Human growth hormo
15	255	98.1	138	9	AA11226	Sequence of protei
16	255	98.1	179	23	AA147922	Human GH-V Seq 10
17	255	98.1	191	7	AA160116	Sequence of human
18	255	98.1	191	18	AA1020110	Protein sequence o
19	255	98.1	191	19	AA171289	Human growth hormo
20	255	98.1	191	20	AA115809	Primary amino acid
21	255	98.1	191	20	AA164196	Natural human 22kd
22	255	98.1	191	20	AA104197	Mutant human 22kd
23	255	98.1	191	21	AA178425	Human growth hormo
24	255	98.1	191	22	AA117445	Human growth hormo
25	255	98.1	191	22	AA117486	Human growth hormo
26	255	98.1	191	23	AA194855	Human growth hormo
27	255	98.1	191	23	AA194856	Human growth hormo
28	255	98.1	191	23	AA194857	Human growth hormo
29	255	98.1	191	23	AA194858	Human growth hormo
30	255	98.1	191	23	AA194859	Human growth hormo
31	255	98.1	191	23	AA194860	Human growth hormo
32	255	98.1	191	23	AA194861	Human growth hormo
33	255	98.1	191	23	AA194862	Human growth hormo
34	255	98.1	191	23	AA194863	Human growth hormo
35	255	98.1	191	23	AA194864	Human growth hormo
36	255	98.1	191	23	AA194865	Human growth hormo
37	255	98.1	191	23	AA194866	Human growth hormo
38	255	98.1	191	23	AA194867	Human growth hormo
39	255	98.1	191	23	AA194868	Human growth hormo
40	255	98.1	191	23	AA194869	Human growth hormo
41	255	98.1	191	23	AA194870	Human growth hormo
42	255	98.1	191	23	AA194871	Human growth hormo
43	255	98.1	191	23	AA194872	Human growth hormo
44	255	98.1	191	23	AA194873	Human growth hormo
45	255	98.1	191	23	AA194874	Human growth hormo

ALIGNMENTS

RESULT 1  
AA142855 standard; protein: 49 AA.  
XX AC AA142855;  
XX DT 19-JAN-2000 (first entry)  
XX DE Human growth hormone (hGH) N terminal fragment B1.  
XX KW Growth hormone; chaperone; intramolecular; insulin; precursor;  
XX KW folding; conformation; chimeric protein; cleavable; recombinant;  
XX KW production; yield.  
XX OS Homo sapiens.  
XX PN WO9950302-A1.  
XX PD 07-OCT-1999.  
XX PF 31-MAR-1998; 98WO-CN00052.  
XX PR 31-MAR-1998; 98WO-CN00052.  
XX PA (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.  
XX PI Gan Z.  
XX DR WPI: 1999-610839/52.  
XX PT New chimeric proteins containing human growth hormone fragment, used particularly for the production of human insulin.

XX PS Claim 4; Page 28; 46pp; English.

XX CC This sequence represents an N-terminal fragment of human growth

XX CC hormone (hGH) which is a component of a chimeric protein,

XX CC hGH-mini-proinsulin (AA42860). The hGH portion of the chimeric protein

XX CC acts as an intramolecular chaperone (IMC) for the insulin precursor,

XX CC enabling it to fold correctly. A cleavable peptide linker with a

XX CC C-terminal Arg residue (AA42857) enables the hGH portion of the

XX CC chimeric protein to be removed after folding has taken place. Production

XX CC of recombinant human insulin via an hGH-proinsulin chimeric protein can

XX CC provide human insulin with correctly linked cysteine bridges with

XX CC fewer necessary procedural steps, and hence resulting in a higher yield

XX CC of human insulin. The IMC sequences not only protect insulin sequences

XX CC from intracellular degradation by a microorganism host, but also promote

XX CC the folding of the fused insulin precursor, facilitate the solubility of

XX CC the fusion protein and decrease the intermolecular interactions among

XX CC the fusion proteins, thus allowing folding of the fused insulin precursor

XX CC at commercially useful high concentrations. The procedural steps of

XX CC cyanogen bromide cleavage, oxidative sulphydrololysis and related

XX CC purification steps can thus be eliminated, along with the use of high

XX CC concentrations of mercaptan or the use of hydrophobic absorbent resins.

XX SQ Sequence 49 AA:

Query Match 100.0%; Score 260; DB 20; Length 49;

Best Local Similarity 100.0%; Pred. No. 2, 3e-25;

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MEPTPLSLFNFNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNP 49

DB 1 MEPTPLSLFNFNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNP 49

## RESULT 2

AA42856

ID AA42856 standard; protein; 92 AA.

XX AC AA42856;

XX DT 19-JAN-2000 (first entry)

XX DE Human growth hormone (hGH) N-terminal fragment #2.

XX KW Growth hormone; chaperone; intramolecular; insulin; precursor;

XX KW folding; conformation; chimeric protein; cleavable; recombinant;

XX KW production; yield.

XX OS Homo sapiens.

XX PN WO9950302-A1.

XX PD 07-OCT-1999.

XX PF 31-MAR-1998; 98WO-CN00052.

XX PR 31-MAR-1998; 98WO-CN00052.

XX PA (TONG) TONGHUA GANTECH BIOTECHNOLOGY LTD.

XX PI Gan Z;

XX WPI: 1999-610839/52.

XX PT New chimeric proteins containing human growth hormone fragment, used

XX PT particularly for the production of human insulin.

XX PS Claim 5; Page 28; 46pp; English.

XX CC This sequence represents an N-terminal fragment of human growth

XX CC hormone (hGH) which is a component of a chimeric protein (AA42861) which also

XX CC contains a human insulin precursor (AA42859). The hGH portion of the

XX CC chimeric protein acts as an intramolecular chaperone (IMC) for the

CC insulin precursor, enabling it to fold correctly. A cleavable peptide

CC linker with a C-terminal Arg residue (AA42857) enables the hGH portion

CC of the chimeric protein to be removed after folding has taken place.

CC Production of recombinant human insulin via an hGH-proinsulin chimeric

CC protein can provide human insulin with correctly linked cysteine bridges

CC with fewer necessary procedural steps, and hence resulting in a higher

CC yield of human insulin. The IMC sequences not only protect insulin

CC sequences from intracellular degradation by a microorganism host, but

CC also promote the folding of the fused insulin precursor, facilitate the

CC solubility of the fusion protein and decrease the intermolecular

CC interactions among the fusion proteins, thus allowing folding of the

CC fused insulin precursor at commercially useful high concentrations. The

CC procedural steps of cyanogen bromide cleavage, oxidative sulphydrololysis

CC and related purification steps can thus be eliminated, along with the

CC use of high concentrations of mercaptan or the use of hydrophobic

XX absorbent resins.

SQ Sequence 92 AA:

Query Match 100.0%; Score 260; DB 20; Length 92;

Best Local Similarity 100.0%; Pred. No. 4, 6e-25;

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MEPTPLSLFNFNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNP 49

DB 1 MEPTPLSLFNFNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNP 49

## RESULT 3

AA42860

ID AA42860 standard; protein; 107 AA.

XX AC AA42860;

XX DT 19-JAN-2000 (first entry)

XX DE hGH-mini-proinsulin chimeric protein.

XX KW Insulin; precursor; growth hormone; chaperone; intramolecular;

XX KW folding; conformation; chimeric protein; cleavable; recombinant;

XX KW production; yield.

XX OS Synthetic.

XX OS Homo sapiens.

XX PN WO9950302-A1.

XX PD 07-OCT-1999.

XX PF 31-MAR-1998; 98WO-CN00052.

XX PR 31-MAR-1998; 98WO-CN00052.

XX PA (TONG) TONGHUA GANTECH BIOTECHNOLOGY LTD.

XX PI Gan Z;

XX WPI: 1999-610839/52.

XX PT New chimeric proteins containing human growth hormone fragment, used

XX PT particularly for the production of human insulin.

XX PS Claim 13; Page 30; 46pp; English.

XX CC This sequence represents a chimeric protein, hGH-mini-proinsulin.

XX CC This chimeric protein contains an N-terminal fragment of human growth

XX CC hormone (hGH) of the sequence given in AA42855, a cleavable peptide

XX CC linker (AA42857), and a human insulin precursor comprising insulin

XX CC A and B chains (AA42859). The hGH portion of the chimeric protein acts

XX CC as an intramolecular chaperone (IMC) for the insulin precursor,

XX CC enabling it to fold correctly. The cleavable peptide linker has a

XX CC C-terminal Arg residue which enables the hGH portion of the

XX CC chimeric protein to be removed after folding has taken place. Production

of recombinant human insulin via an hGH-proinsulin chimeric protein can provide human insulin with correctly linked cysteine bridges with fewer necessary procedural steps, and hence resulting in a higher yield of human insulin. The IMC sequences not only protect insulin sequences from intracellular degradation by a microorganism host, but also promote the folding of the fused insulin precursor, facilitate the solubility of the fusion protein and decrease the intermolecular interactions among the fusion proteins, thus allowing folding of the fused insulin precursor at commercially useful high concentrations. The procedural steps of cyanoogen bromide cleavage, oxidative sulphydrololysis and related purification steps can thus be eliminated, along with the use of high concentrations of mercaptan or the use of hydrophobic absorbent resins.

Query Match 100.0%; Score 260; DB 20; Length 107;  
Best Local Similarity 100.0%; Pred. No. 54e-25;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MFPTPLSLRFLNAMLRAHRLHJLAFTYQEFEEAYIPKQKYSFQNP 49  
DB 1 MFPTPLSLRFLNAMLRAHRLHJLAFTYQEFEEAYIPKQKYSFQNP 49

RESULT 4  
AAW92265  
ID AAW92265 standard; Protein: 134 AA.

AAW92265;  
08-JUN-1999 (first entry)  
Human anti-angiogenic peptide 16K hGH Met-1pro133.  
Human; anti-angiogenic; prolactin; placental lactogen; hPL; angiogenesis; growth hormone; hGH; hGH-V; capillary endothelial cell proliferation; placental vascularisation; pregnancy; treatment; angiogenic disease; tumour; inhibitor; malignant; angiofibroma; arteriovenous malformation; arthritis; atherosclerotic plaques; corneal graft neovascularisation; wound healing; proliferative retinopathy; macular degeneration; trachoma; granulation; glaucoma; ocular; uveitis; culture; Osler-Weber syndrome; psoriasis; fibroplasia; scleroderma; Kaposi's sarcoma; vascular adhesion; ulcer; leukaemia; reproductive disorder; contraceptive agent; gene therapy; pre-eclampsia; intrauterine growth retardation; placental dysfunction.

Homo sapiens.  
W00451423 A1.  
19 NOV 1998.  
12 MAY 1998; 96WO-US09691.  
13-MAY-1997; 97US-0046394.  
(REPT) UNIV CALIFORNIA.  
Marshall JA, Struman I, Taylor K, Weiner Rf;  
WPI: 1999-045192/04.  
N-PSOR; AAK31707.  
New anti-angiogenic peptides - comprise N-terminal fragments of human placental lactogen, human growth hormone, growth hormone variant or human prolactin

Claim 4: Page 49-50; 87pp; English.  
This invention describes novel human anti-angiogenic peptides derived from 10 to 150 consecutive amino acids selected from the N-terminal end of human placental lactogen (hPL), human growth hormone (hGH), growth hormone variant (hGH-V), or human prolactin. Such peptides (i) inhibit

capillary endothelial cell proliferation and organisation (ii) inhibit angiogenesis in chick chorioallantoic membrane and (iii) binds to at least one specific receptor which does not bind an intact full length hGH, hPL, prolactin or hGH-V. The invention also describes a method for diagnosing a probable abnormality of placental vascularisation during pregnancy. The peptides can be used for treating an angiogenic disease in a subject, for inhibiting tumour formation or growth in a patient, or for modulating vascularisation of a patient's placenta. In particular, the peptides can be used for preventing or treating e.g. malignant tumours, angiofibroma, arteriovenous malformation, arthritic such as rheumatoid arthritis, atherosclerotic plaques, corneal graft neovascularisation, delayed wound healing, proliferative retinopathy such as diabetic retinopathy, macular degeneration, granulation such as those occurring in haemophiliac joints, inappropriate vascularisation in wound healing such as hypertrophic scars or keloid scars, neovascular glaucoma, ocular tumour, uveitis, non-union fractures, Osler-Weber syndrome, psoriasis, pyogenic glaucoma, retrolental fibroplasia, scleroderma, solid tumours, Kaposi's sarcoma, trachoma, vascular adhesions, chronic varicose ulcers, leukaemia, and reproductive disorders such as follicular and luteal cysts and chorioepithelioma. They can also be used as contraceptive agents. DNA encoding the peptides can be used in gene therapy. The measurement of choroidal levels of N-terminal fragments of hPL, hGH-V, prolactin or hPL can be used in assays for impairment of vascular development associated with pre-eclampsia, intrauterine growth retardation, and placental dysfunction.

Sequence 134 AA;

Query Match 100.0%; Score 260; DB 20; Length 134;  
Best Local Similarity 100.0%; Pred. No. 7e-25;  
Matches 49; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 MFPTPLSLRFLNAMLRAHRLHJLAFTYQEFEEAYIPKQKYSFQNP 49  
DB 1 MFPTPLSLRFLNAMLRAHRLHJLAFTYQEFEEAYIPKQKYSFQNP 49

RESULT 5  
AAP91041  
ID AAP91041 standard; Protein: 140 AA.  
AAP91041;  
14-DEC-1989 (first entry).  
Human growth hormone segment.  
Human growth hormone; fusion protein; thrombin; geriatric dementia; nervous disorders; human nerve factor.

Homo sapiens (human).  
EP329175-A.  
23-AUG-1989.  
17-FEB-1989; 89EP-0102795.  
19-FEB-1988; 88JP-0035042.  
(TOYU) TOSOH CORP.  
Ohtsuka F;  
WPI: 1989-243092/34.  
New human nerve growth factor gene encoding fusion protein - having cleavage site for thrombin, useful for treating geriatric dementia, etc.  
Disclosure: page 21; 38pp; English.  
Human growth hormone segment, used at the N terminal of a fusion

CC protein, which contains a thrombin recognition site, and human beta nerve  
 CC growth factor (beta-NGF) at the C-terminal. Beta-NGF can be used to  
 CC control geriatric dementia and other nervous disorders, and can be  
 CC released from the fusion protein by incubation with thrombin (see  
 CC AAN90577-8, AAP91034, AAP91299).

XX SQ Sequence 140 AA;

Query Match 100.0%; Score 260; DB 10; Length 140;  
 Best Local Similarity 100.0%; Pred. No. 7.9e-25;  
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEPTPLSLRFDNMLRAHRLHQLAFDTYQFEFEAYIPKEOKYSFLONP 49  
 |||||  
 Db 1 MEPTPLSLRFDNMLRAHRLHQLAFDTYQFEFEAYIPKEOKYSFLONP 49

# RESULT 6

AAV42861  
 ID AAV42861 standard; protein: 150 AA.

XX AC AAV42861;

DT 19 JAN-2000 (first entry)

XX Chimeric protein, SEQ ID 7.

XX Insulin; precursor: growth hormone; chap- o; intramolecular;  
 KW folding; conformation; chimeric protein; available; recombinant;  
 KW production; yield.

XX Synthetic.

OS Homo sapiens.

XX WO950302-A1.

XX 07-JUN-1999.

XX 31 MAR-1998; 98WO-CN00052.

XX 31 MAR-1998; 98WO-CN00052.

XX (TUNGHUA GANTECH BIOTECHNOLOGY LTD.

XX Gan Z;

XX WPI: 1999.610839/52.

XX New chimeric proteins containing human growth hormone fragment, used  
 PT particularly for the production of human insulin.

XX Claim 14; Page 30-31; 46pp; English.

XX This sequence represents a chimeric protein, which contains an  
 CC N-terminal fragment of human growth hormone (hGH) of the sequence given  
 CC in AAY42856, a cleavable peptide linker (AAY42857), and a human insulin  
 CC precursor comprising insulin A and B chains (AAY42859). The hGH portion  
 CC of the chimeric protein acts as an intramolecular chaperone (IMC) for  
 CC the insulin precursor, enabling it to fold correctly. The cleavable  
 CC peptide linker has a C-terminal Arg residue which enables the hGH portion  
 CC of the chimeric protein to be removed after folding has taken place.  
 CC Production of recombinant human insulin via an hGH-proinsulin chimeric  
 CC protein can provide human insulin with correctly linked cysteine bridges  
 CC with fewer necessary procedural steps, and hence resulting in a higher  
 CC yield of human insulin. The IMC sequences not only protect insulin  
 CC sequences from intracellular degradation by a microorganism host, but  
 CC also promote the folding of the fused insulin precursor, facilitate the  
 CC solubility of the fusion protein and decrease the intermolecular  
 CC interactions among the fusion proteins, thus allowing folding of the  
 CC fused insulin precursor at commercially useful high concentrations. The  
 CC procedural steps of cyanogen bromide cleavage, oxidative sulphytolysis  
 CC and related purification steps can thus be eliminated, along with the use  
 CC of high concentrations of mercaptan or the use of hydrophobic absorbent

CC resins.

XX SQ Sequence 150 AA;

Query Match 100.0%; Score 260; DB 20; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 7.9e-25;  
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEPTPLSLRFDNMLRAHRLHQLAFDTYQFEFEAYIPKEOKYSFLONP 49  
 |||||  
 Db 1 MEPTPLSLRFDNMLRAHRLHQLAFDTYQFEFEAYIPKEOKYSFLONP 49

# RESULT 7

AAP90129  
 ID AAP90129 standard; protein: 192 AA.

XX AC AAP90129;

XX 25 MAR-2004 (updated)

DT 06-FEB-1996 (revised)

DT 01-NOV-1989 (first entry)

XX Human growth hormone.

XX Human growth hormone; fusion protein; recombinant

XX Home sapiens (Human).

XX JP01144981-A.

XX 07-JUN 1989.

XX 02-JEC-1987; 87JP-0304937.

XX 02-DEC-1987; 87JP-0304937.

XX (WAKT) WAKUNAGA SEIVAKU KK.

XX WPI: 1989-209284/29.

XX N-PSDB: AAN90265.

XX Recombinant vector contg. fusion protein consisting of human  
 PT growth hormone or deriv. ligated to foreign protein, for stability  
 PT and high yield.  
 XX Disclosure: Fig 1: 15pp; Japanese.

XX The invention consists of a vector contg. a fusion protein which is  
 CC formed by ligating, downstream of a promoter, hGH or a deriv. (pref.  
 CC formed by substn. of Met-14 with leu) and a foreign protein.  
 CC Stability of the vector in the host is greatly increased so the  
 CC protein yield is higher.  
 CC (Updated on 25-MAR-2003 to correct PA field.)

XX SQ Sequence 192 AA;

Query Match 100.0%; Score 260; DB 10; Length 192;  
 Best Local Similarity 100.0%; Pred. No. 1e-24;  
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEPTPLSLRFDNMLRAHRLHQLAFDTYQFEFEAYIPKEOKYSFLONP 49  
 |||||  
 Db 1 MEPTPLSLRFDNMLRAHRLHQLAFDTYQFEFEAYIPKEOKYSFLONP 49

# RESULT 8

AAW92264  
 ID AAW92264 standard; protein: 192 AA.

XX AC AAW92264;



DT 08-JUN-1999 (first entry)  
 XX Human anti-angiogenic peptide hGH Met-1Phe191.  
 DE Human: anti-angiogenic; prolactin; placental lactogen; hPL; angiogenesis;  
 KW growth hormone; hGH; hGH-V; capillary endothelial cell proliferation;  
 KW placental vasculature; pregnancy; treatment; angiogenic disease;  
 KW tumour; inhibitor; malignant; angiofibroma; arteriovenous malformation;  
 KW arthritis; atherosclerotic plaques; corneal graft neovascularisation;  
 KW wound healing; proliferative retinopathy; macular degeneration; trachoma;  
 KW granuloma; glaucoma; ocular; uveitis; fracture; Osler-Weber syndrome;  
 KW psoriasis; fibroplasia; scleroderma; Kaposi's sarcoma; vascular adhesion;  
 KW ulcer; leukaemia; reproductive disorder; contraceptive agent;  
 KW gene therapy; pre-eclampsia; intrauterine growth retardation;  
 KW placental dysfunction.  
 XX Homo sapiens.  
 OS  
 XX  
 XX  
 PN 84451323-A1.  
 XX  
 PD 14 NOV-1998.  
 XX  
 PF 12-MAY-1998. 98W/0506651.  
 XX  
 PR 11 MAY 1997. 97US-0046394.  
 XX  
 PA (RESC) UNIV CALIFORNIA.  
 XX  
 XX Martial JA, Struman I, Taylor R, Wainor RI.  
 PI WPI: 1999-045192/04.  
 DR N-PSDB: AAX01706.  
 DR  
 XX  
 XX New anti-angiogenic peptides - comprise N-terminal fragments of  
 PT human placental lactogen, human growth hormone, growth hormone  
 PT variant or human prolactin  
 XX  
 XX Example 3: Page 49: 87pp: English.  
 PS  
 CC This invention describes novel human anti-angiogenic peptides derived  
 CC from 10 to 150 consecutive amino acids selected from the N-terminal end  
 CC of human placental lactogen (hPL), human growth hormone (hGH), growth  
 CC hormone variant (hGH-V), or human prolactin. Such peptides (i) inhibit  
 CC capillary endothelial cell proliferation and (ii) inhibit  
 CC angiogenesis in chick chorioallantoic membrane and (iii) binds to at  
 CC least one specific receptor which does not bind an intact full length  
 CC hGH, hPL, prolactin or hGH-V. The invention also describes a method for  
 CC diagnosing a probable abnormality of placental vascularisation during  
 CC pregnancy. The peptides can be used for treating an angiogenic disease in  
 CC a subject, for inhibiting tumour formation or growth in a patient or for  
 CC modulating vascularisation of a patient's placenta. In particular, the  
 CC peptides can be used for preventing or treating e.g. malignant tumours,  
 CC angiofibroma, arteriovenous malformation, arthritic such as rheumatoid  
 CC arthritis, atherosclerotic plaques, corneal graft neovascularisation,  
 CC delayed wound healing, proliferative retinopathy such as diabetic  
 CC retinopathy, macular degeneration, granulations such as those occurring  
 CC in haemophilic joints, inappropriate vascularisation in wound healing  
 CC such as hypertrophic scars or keloid scars, neovascular glaucoma, ocular  
 CC tumour, uveitis, non-union fractures, Osler-Weber syndrome, psoriasis,  
 CC pyogenic glaucoma, retrolental fibroplasia, scleroderma, solid tumours,  
 CC Kaposi's sarcoma, trachoma, vascular adhesions, chronic varicose ulcers,  
 CC leukaemia, and reproductive disorders such as follicular and luteal cysts  
 CC and chorioncarinoma. They can also be used as contraceptive agents. DNA  
 CC encoding the peptides can be used in gene therapy. The measurement of  
 CC abnormal levels of N-terminal fragments of: hGH, hGH-V, prolactin or hPL  
 CC can be used in assays for impairment of vascular development associated  
 CC with pre-eclampsia, intrauterine growth retardation, and placental  
 CC dysfunction.  
 XX  
 SQ Sequence 192 AA:  
 Query Match 100.0%; Score 260; DB 20; Length 192;  
 Best Local Similarity 100.0%; Pred. No. 1e-24;  
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MFPTIPLSRFLDNAMLAHRLHQLAFDTYQFEFFAYIPKEOKYSFLQNP 49  
 DB 1 MFPTIPLSRFLDNAMLAHRLHQLAFDTYQFEFFAYIPKEOKYSFLQNP 49  
 RESULT 9  
 AAP91299  
 ID AAP91299 standard; protein: 261 AA.  
 AC AAP91299;  
 XX  
 XX 14-DEC-1989 (first entry).  
 DT Human nerve growth factor and human growth hormone fusion protein.  
 DE Human nerve growth factor; fusion protein; thrombin;  
 KW geriatric dementia; nervous disorders; human growth hormone.  
 KW Homo sapiens (human).  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FT Region 1..140  
 FT Region 141..143  
 FT Region 144..261  
 PN EP329175-A.  
 PD 23-AUG-1989.  
 XX  
 PF 17-FEB-1989; 89EP-0102745.  
 XX  
 PR 19-FEB-1988; 88JP-0035042.  
 XX  
 PA (TOYJ) TOSOH CORP.  
 XX  
 XX Ohtsuka E;  
 PI  
 DR WPI: 1989-243092/34.  
 XX  
 PT New human nerve growth factor gene encoding fusion protein  
 PT - having cleavage site for thrombin, useful for treating geriatric  
 PT dementia, etc.  
 XX  
 PS Claim 36: page 31-32: 38pp: English.  
 XX  
 CC Fusion protein consisting of human growth hormone at the  
 CC N-terminal end (1st region), a 3 amino acid sequence representing  
 CC thrombin recognition site, and human beta nerve growth factor (beta-NGF)  
 CC at the C-terminal. Beta-NGF can be used to control geriatric dementia  
 CC and other nervous disorders, and can be released from the fusion  
 CC protein by incubation with thrombin (see AAN50577-8, AAP91034,  
 CC AAP91041).  
 XX  
 SQ Sequence 261 AA:  
 Query Match 100.0%; Score 260; DB 10; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-24;  
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MFPTIPLSRFLDNAMLAHRLHQLAFDTYQFEFFAYIPKEOKYSFLQNP 49  
 DB 1 MFPTIPLSRFLDNAMLAHRLHQLAFDTYQFEFFAYIPKEOKYSFLQNP 49  
 RESULT 10  
 AAP61033  
 ID AAP61033 standard; protein: 272 AA.  
 AC AAP61033;  
 XX  
 XX 25-OCT-1991 (first entry)  
 DT



## RESULT 13

AAR05313  
ID AAR05313 standard; protein: 144 AA.

XX AC AAR05313;

XX DT 19-JUL-1990 (first entry)

XX DE Segment of B-cell stimulatory factor-2 (IL-5).

XX KW B-cell stimulatory factor-2; interleukin-5.

XX OS Homo sapiens.

XX PN J062013375-A.

XX PD 17-JAN-1990.

XX PF 01-JUL 1988; 88JP-0162556.

XX PR 01-JUL 1988; 88JP-0162556.

XX PA (TOYO ) TOYO SODA CORP.

XX WP: 1990-052207/09.

XX N PSDB; AAQ02028.

XX PT Propn. of human B-cell differentiation factor - from specified DNA  
sequence segment, by recombinant DNA technique, gives protein of  
specified amino acid sequence.

XX PS Disclosure: Page 9; 17pp; Japanese.

XX CC The sequence encoding this protein can be fused with DNA encoding B-cell  
differentiation factor (IL-6) and ligated into an expression vector for  
production of a fusion protein.  
See also AAR05311.

XX SQ Sequence 144 AA:

Query Match 98.8%; Score 257; DB 11; Length 144;

Best Local Similarity 98.0%; Pred. No. 18e-24;

Matches 48; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEPTPLSLRFDNAMLRAHRLHQLAFDTYQEFEEAIPKEQKYSFLQNP 49

DB 1 MEPTPLSLRFDNAMLRAHRLHQLAFDTYQEFEEAIPKEQKYSFLQNP 49

## RESULT 14

ABB77327  
ID ABB77327 standard; Protein: 204 AA.

XX AC ABB77327;

XX DT 17-JUN-2002 (first entry)

XX DE Human growth hormone with collagenase recognition site.

XX KW Human; growth hormone; collagenase; recognition site.

XX OS Homo sapiens.

XX PN KR289691-A.

XX PD 15-MAY-2001.

XX PF 24-DEC-1993; 93KR-0030318.

XX PR 24-DEC-1993; 93KR-0030318.

XX PA (GLDS ) LG CHEM LTD.

PI Yoo JG, Song YH;

XX DR WPI: 2002-185396/24.

XX DR N-PSDB; AHL55999.

XX PT Recombinant human growth hormone having collagenase recognition region -

XX PS Disclosure: Fig 3; 8pp; Korean.

XX CC The invention relates to recombinant human growth hormone having a  
collagenase recognition region.

XX SQ Sequence 204 AA:

Query Match 98.5%; Score 256; DB 23; Length 204;

Best Local Similarity 98.0%; Pred. No. 3.5e-24;

Matches 48; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEPTPLSLRFDNAMLRAHRLHQLAFDTYQEFEEAIPKEQKYSFLQNP 49

DB 1 MEPTPLSLRFDNAMLRAHRLHQLAFDTYQEFEEAIPKEQKYSFLQNP 61

## RESULT 15

AAP81226  
ID AAP81226 standard; protein: 138 AA.

XX AC AAP81226;

XX DT 25-MAR-2003 (updated)

XX DT 20-NOV-1990 (first entry)

XX DE Sequence of protein with somatomedin-like activity.

XX KW Growth hormone.

XX OS Synthetic.

XX PN JP63167798-A.

XX PD 11-JUL-1988.

XX PF 29-DEC-1986; 86JP-0310177.

XX PR 29-DEC-1986; 86JP-0310177.

XX PA (TOYO ) TOYO SODA MFG CO LTD.

XX WP: 1988-232632/33.

XX N-PSDB; AAP81605.

XX PT Polypeptide with somatomedin-like activity -  
by culturing bacterium transformed by plasmid contg. gene  
segment with specified DNA sequence

XX PS Claim 2(1); Page 609; 9pp; Japanese.

XX CC The polypeptide (AAP81226) with somatomedin-like activity and the DNA  
(AAP81605) encoding it are claimed. A Met residue gp. may be added to  
the N-terminal. The polypeptide acts on the bone structure of mammals,  
including humans, to promote bone growth. The polypeptide has high  
production rate and is easily extracted from bacterial culture medium  
and refined for use as a bone growth accelerator.  
(updated on 25-MAR-2003 to correct PA field.)

XX SQ Sequence 138 AA:

Query Match 98.1%; Score 255; DB 9; Length 138;

Best Local Similarity 100.0%; Pred. No. 3e-24;

Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTPLSLRFDNAMLRAHRLHQLAFDTYQEFEEAIPKEQKYSFLQNP 49

DB 2 FPTPLSLRFDNAMLRAHRLHQLAFDTYQEFEEAIPKEQKYSFLQNP 49

DD 1 FPTIPLSLFONAMLRHRLHQLAFDTVOFFEEAYIPKEQKYSFLQNP 48

Search completed: September 15, 2003, 12:00:56  
Job time : 35.4229 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:57:40 : Search time 11.2401 Seconds  
(without alignments)  
184.449 Million cell updates/sec

Title: US-09-423-100-1

Perfect score: 260

Sequence: 1 MEPTPLSLRFLONAMLRARR.....OFFEEAYIPKEQKYSFLQNP 49

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 428717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 428717

Minimum db seq length: 0

Maximum db seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2.6/plodata/1/iaa/5A MB.pep:\*  
2: /cgn2.6/plodata/1/iaa/5B MB.pep:\*  
3: /cgn2.6/plodata/1/iaa/6A MB.pep:\*  
4: /cgn2.6/plodata/1/iaa/6B MB.pep:\*  
5: /cgn2.6/plodata/1/iaa/PCMB.pep:\*  
6: /cgn2.6/plodata/1/iaa/backtilles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	260	100.0	192	1	US-08-093-383-1
2	255	98.1	191	4	US-09-284-878-5
3	255	98.1	194	2	US-08-383-621-4
4	255	98.1	194	3	US-08-459-906-4
5	255	98.1	217	3	US-08-589-028-10
6	255	98.1	217	3	US-08-784-582-10
7	255	98.1	217	3	US-08-785-271-10
8	255	98.1	217	3	US-08-759-628-11
9	255	98.1	217	4	US-09-284-878-1
10	255	98.1	241	4	US-09-424-620B-25
11	255	98.1	245	4	US-09-280-030-66
12	255	98.1	274	3	US-08-784-582-71
13	255	98.1	360	3	US-08-784-582-73
14	249	95.8	191	4	US-09-465-461-1
15	249	95.8	217	1	US-08-187-756C-4
16	249	95.8	217	1	US-08-469-486-51
17	249	95.8	217	2	US-08-469-651-51
18	249	95.8	217	2	US-08-710-317-4
19	249	95.8	217	4	US-09-411-657-4
20	248	95.4	191	3	US-08-800-215C-16
21	248	95.4	191	3	US-08-800-215C-18
22	248	95.4	191	3	US-08-800-215C-20
23	248	95.4	400	4	US-09-420-819-17
24	248	95.4	401	4	US-09-420-819-16
25	237	91.2	71	1	US-08-314-586-24
26	233	89.6	70	1	US-07-920-519-24
27	233	89.6	70	3	US-08-115-753-26

28	164.5	63.3	191	1	US-08-458-824-8	Sequence 8, Appli
29	164	63.1	176	3	US-08-791-728-1	Sequence 1, Appli
30	164	63.1	176	4	US-08-990-774-1	Sequence 1, Appli
31	161.5	62.1	191	1	US-07-963-311D-4	Sequence 4, Appli
32	159.5	61.3	190	1	US-08-388-267C-2	Sequence 2, Appli
33	159.5	61.3	190	4	US-09-277-720-2	Sequence 2, Appli
34	159.5	61.3	191	6	5210180-1	Patent No. 5210180
35	159.5	61.3	193	1	US-07-621-197C-2	Sequence 2, Appli
36	159.5	61.3	193	1	US-08-363-982-2	Sequence 1, Appli
37	159.5	61.3	193	3	US-08-381-621-1	Sequence 1, Appli
38	159.5	61.3	193	3	US-08-459-906-1	Sequence 1, Appli
39	159.5	61.3	216	2	US-09-105-651-1	Sequence 1, Appli
40	159.5	61.3	216	2	US-09-105-651-3	Sequence 3, Appli
41	158	60.8	176	3	US-08-791-728-2	Sequence 2, Appli
42	158	60.8	176	4	US-08-990-774-2	Sequence 2, Appli
43	158	60.8	177	1	US-08-187-756C-6	Sequence 6, Appli
44	158	60.8	177	2	US-08-710-324A-6	Sequence 6, Appli
45	158	60.8	177	4	US-09-411-657-5	Sequence 6, Appli

#### ALIGNMENTS

RESULT 1  
US-08-093-383-1  
: Sequence 1, Application US/08093383  
: Patent No. 5489529  
: GENERAL INFORMATION:  
: APPLICANT: DeBoer, Herman A.  
: APPLICANT: Heyncker, Herbert L.  
: TITLE OF INVENTION: DNA for Expression of Bovine Growth Hormone  
: NUMBER OF SEQUENCES: 30  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Genentech, Inc.  
: STREET: 460 Point San Bruno Blvd  
: CITY: South San Francisco  
: STATE: California  
: COUNTRY: USA  
: ZIP: 94080  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: patin (Genentech)  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/093.383  
: FILING DATE: 14-JUL-1993  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 07/619827  
: FILING DATE: 28-NOV-1990  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 07/198824  
: FILING DATE: 05-APR 1988  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 06/632361  
: FILING DATE: 19-JUL-1984  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 06/101687  
: FILING DATE: 18-SEP-1981  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Johnston, Sean A.  
: REGISTRATION NUMBER: P35,910  
: REFERENCE/DOCKET NUMBER: 46C4  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 415/225-3562  
: TELEFAX: 415/952-9881  
: TELEX: 910/371-7168  
: INFORMATION FOR SEQ ID NO: 1:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 192 amino acids  
: TYPE: amino acid

```

; TOPOLOGY: linear
; US-08-053-383-1

Query Match      100.0%  Score 260;  DB 1;  Length 192;
Best Local Similarity 100.0%;  Pred. No. 4,6e-30;
Matches 49;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY  1 MFPTIPLSRFDNAMLRAHRLHQLAFDTYQFFEEAYIPKEOKYSFLQNP 49
      |||||||
DB  1 MFPTIPLSRFDNAMLRAHRLHQLAFDTYQFFEEAYIPKEOKYSFLQNP 49

RESULT 4
US-09-284-878-5
; Sequence 5, Application US/09284878
; Patent No. 6342175
; GENERAL INFORMATION:
; APPLICANT: Olazaran, Martha Guerrero
; APPLICANT: Saldaña, Hugo Parrota
; APPLICANT: Salvado, Jose Maria Viader
; TITLE OF INVENTION: Genetically Modified Methylotrophic P. pastoris Yeast for the
; TITLE OF INVENTION: Production and Secretion of the Human Growth Hormone
; FILE REFERENCE: 1829-0010036
; CURRENT APPLICATION NUMBER: US/09/284.878
; CURRENT FILING DATE: 1999-07-21
; PRIOR APPLICATION NUMBER: PCT/MX97/00033
; PRIOR FILING DATE: 1997-10-24
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-284-878-5

Query Match      98.18;  Score 255;  DB 4;  Length 191;
Best Local Similarity 100.0%;  Pred. No. 2.4e-29;
Matches 48;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY  2 FPTIPLSKLFDNAMLRAHRLHQLAFDTYQFFEEAYIPKEOKYSFLQNP 49
      |||||||
DB  1 FPTIPLSKLFDNAMLRAHRLHQLAFDTYQFFEEAYIPKEOKYSFLQNP 48

RESULT 4
US-08-459-906-4
; Sequence 4, Application US/08459406
; Patent No. 5951972
; GENERAL INFORMATION:
; APPLICANT: Daley, Michael J.
; APPLICANT: Buckwalter, Brian L.
; APPLICANT: Cady, Susan M.
; APPLICANT: Shieh, Hong-Ming
; APPLICANT: Bohlen, Peter
; APPLICANT: Seddon, Andrew P.
; TITLE OF INVENTION: Stabilization of Somatotropins and Other
; TITLE OF INVENTION: Proteins by Modification of Cysteine Residues
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07470-8426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,906
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Darryl L.
; REGISTRATION NUMBER: 34,276
; REFERENCE/DOCKET NUMBER: 31,278-03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-831-3247
; TELEFAX: 201-831-3305
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 194 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-459-906-4

```

```

; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/766,142
; FILING DATE: 25-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsevdos, Estelle J.
; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,278-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 203-321-2756
; TELEFAX: 203-321-2971
; TELEX: 203-710-474-4059
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 194 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-383-621-4

Query Match      98.18;  Score 255;  DB 4;  Length 194;
Best Local Similarity 100.0%;  Pred. No. 2.5e-29;
Matches 49;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY  2 FPTIPLSKLFDNAMLRAHRLHQLAFDTYQFFEEAYIPKEOKYSFLQNP 49
      |||||||
DB  4 FPTIPLSKLFDNAMLRAHRLHQLAFDTYQFFEEAYIPKEOKYSFLQNP 51

RESULT 4
US-08-459-906-4
; Sequence 4, Application US/08459406
; Patent No. 6010995
; GENERAL INFORMATION:
; APPLICANT: Daley, Michael J.
; APPLICANT: Buckwalter, Brian L.
; APPLICANT: Cady, Susan M.
; APPLICANT: Shieh, Hong-Ming
; APPLICANT: Bohlen, Peter
; APPLICANT: Seddon, Andrew P.
; TITLE OF INVENTION: Stabilization of Somatotropins and Other
; TITLE OF INVENTION: Proteins by Modification of Cysteine Residues
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07470-8426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,906
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Darryl L.
; REGISTRATION NUMBER: 34,276
; REFERENCE/DOCKET NUMBER: 31,278-03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-831-3247
; TELEFAX: 201-831-3305
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 194 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-459-906-4

```

Query Match 98.1% Score 255; DB 3; Length 194;  
Best Local Similarity 100.0%; Pred. No. 2.5e-29;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNMLRAHRLHQLAFDTYQEEFEAYIPKEQKYSFLQNP 49  
|||||  
DB 4 FPTPLSRFDNMLRAHRLHQLAFDTYQEEFEAYIPKEQKYSFLQNP 51

## RESULT 5

US-08-589-028-10  
; Sequence 10, Application US/08589028  
; Patent No. 6087129  
; GENERAL INFORMATION:  
; APPLICANT: Newgard, Christopher B.  
; APPLICANT: Halban, Philippe  
; APPLICANT: No. 6087129mington, Karl D.  
; APPLICANT: Clark, Samuel A.  
; APPLICANT: Thiipen, Anice E.  
; APPLICANT: Quade, Christian  
; APPLICANT: Kruse, Fred  
; TITLE OF INVENTION: Recombinant Expression of Proteins from  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P. O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77210-4433

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/589,028  
FILING DATE: Concurrently herewith  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Highlander, Steven L.  
REGISTRATION NUMBER: 47,642

REFERENCE/DOCKET INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (512) 474-7577

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:  
LENGTH: 217 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear

US-08-589-028-10

Query Match 98.1% Score 255; DB 3; Length 217;  
Best Local Similarity 100.0%; Pred. No. 2.5e-29;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNMLRAHRLHQLAFDTYQEEFEAYIPKEQKYSFLQNP 49  
|||||  
DB 27 FPTPLSRFDNMLRAHRLHQLAFDTYQEEFEAYIPKEQKYSFLQNP 74

## RESULT 6

US-08-744-582-10  
; Sequence 10, Application US/08744582  
; Patent No. 6110707  
; GENERAL INFORMATION:  
; APPLICANT: Newgard, Christopher B.  
; APPLICANT: Halban, Philippe A.  
; APPLICANT: No. 6110707mington, Karl D.

APPLICANT: Clark, Samuel A.  
APPLICANT: Thiipen, Anice E.  
APPLICANT: Quade, Christian  
APPLICANT: Kruse, Fred  
APPLICANT: McGarry, Dennis  
TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM  
TITLE OF INVENTION: SECRETORY CELL LINES  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P. O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/784,582  
FILING DATE: Concurrently herewith  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/028,427  
FILING DATE: 15-OCT-1996

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/389,028  
FILING DATE: 19-JAN-1996

ATTORNEY/AGENT INFORMATION:  
NAME: Highlander, Steven L.  
REGISTRATION NUMBER: 37,642

REFERENCE/DOCKET NUMBER: UTSO-514  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 512/418-3000

TELEFAX: 512/474-7577  
INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:  
LENGTH: 217 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear

US-08-784-582-10

Query Match 98.1% Score 255; DB 3; Length 217;  
Best Local Similarity 100.0%; Pred. No. 2.5e-29;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNMLRAHRLHQLAFDTYQEEFEAYIPKEQKYSFLQNP 49  
|||||  
DB 27 FPTPLSRFDNMLRAHRLHQLAFDTYQEEFEAYIPKEQKYSFLQNP 74

## RESULT 7

US-08-785-271-10  
; Sequence 10, Application US/08785271  
; Patent No. 6194176  
; GENERAL INFORMATION:  
; APPLICANT: Newgard, Christopher B.  
; APPLICANT: Halban, Philippe A.  
; APPLICANT: No. 6194176mington, Karl D.  
; APPLICANT: Clark, Samuel A.  
; APPLICANT: Thiipen, Anice E.  
; APPLICANT: Quade, Christian  
; APPLICANT: Kruse, Fred  
; TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM  
; TITLE OF INVENTION: SECRETORY CELL LINES  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P. O. Box 4433  
; CITY: Houston

```

STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/785,271
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/589,028
FILING DATE: 19-1AN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: 615,513
TELEPHONE: 512/418-3003
TELEFAX: 512/434-7577
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 217 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-785 271-10

Query Match 98.1% Score 255; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 2.9e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFFEEAYIPKEOKYSFLGNP 49
DB 27 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFFEEAYIPKEOKYSFLGNP 74

RESULT 9
US-08-785 628-11
Sequence 1; Application US/08/59628
Patent No. 6225446
GENERAL INFORMATION:
APPLICANT: Altman, Scott W.
APPLICANT: Rock, Fernando L.
APPLICANT: Bazan, J. Fernando
APPLICANT: Kastelein, Robert A.
TITLE OF INVENTION: MUTATIONAL VARIANTS OF MAMMALIAN PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: INAX Research Institute
STREET: 901 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/759,628
FILING DATE: 05-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/008,574
FILING DATE: 06-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ching, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: DX05520

```

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-852-9196
TELEFAX: 415-496-1200
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 217 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Peptide
LOCATION: 32..53
FEATURE:
NAME/KEY: Peptide
LOCATION: 94..115
FEATURE:
NAME/KEY: Peptide
LOCATION: 133..154
FEATURE:
NAME/KEY: Peptide
LOCATION: 192..216
OTHER INFORMATION: /note "The peptides above are
OTHER INFORMATION: depicted in Figure 1."
US 08-759-628 11

Query Match 98.1% Score 255; DB 4; Length 217;
Best Local Similarity 100.0%; Pred. No. 2.9e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFFEEAYIPKEOKYSFLGNP 49
DB 27 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFFEEAYIPKEOKYSFLGNP 74

RESULT 9
US-09-284-878-1
Sequence 1; Application US/09284878
Patent No. 6342375
GENERAL INFORMATION:
APPLICANT: Olazaran, Martha Guerrero
APPLICANT: Saldana, Hugo Barrera
APPLICANT: Salvado, Jose Maria Vinder
TITLE OF INVENTION: Genetically Modified Methylophilic P. pastoris Yeast for the
Production and Secretion of the Human Growth Hormone
FILE REFERENCE: 1829.0010000
CURRENT APPLICATION NUMBER: US/09/284,878
CURRENT FILING DATE: 1999-07-21
PRIOR APPLICATION NUMBER: PCT/MX97/06034
PRIOR FILING DATE: 1997-10-24
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 217
TYPE: PRI
ORGANISM: Homo sapiens
US-09-284-878-1

Query Match 98.1% Score 255; DB 4; Length 217;
Best Local Similarity 100.0%; Pred. No. 2.9e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFFEEAYIPKEOKYSFLGNP 49
DB 27 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFFEEAYIPKEOKYSFLGNP 74

RESULT 10
US 09-424-620R-25
Sequence 25; Application US/09424620R
Patent No. 6391585
GENERAL INFORMATION:
APPLICANT: HANIL SYNTHETIC FIBER CO., LTD.

```



JANG, Ki-Ryong  
MOON, Jae-Woong  
BAE, Cheon-Soon  
YANG, Doo-Suk  
LEE, Jee-Won  
SFONG, Baik-Lin

TITLE OF INVENTION: Process for preparing recombinant proteins using highly efficient expression vector from *Saccharomyces cerevisiae*

NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BACHMAN & LAPOINTE, P.C.  
STREET: Suite 1201, 900 Chapel Street  
CITY: New Haven  
STATE: Connecticut  
COUNTRY: U.S.A.  
ZIP: 06510-2802

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB storage  
COMPUTER: IBM  
OPERATING SYSTEM: WINDOWS 95/NT

CURRENT APPLICATION DATA:  
SOFTWARE: MS WORD  
FILING DATE: 24-NOV-99  
FILING DATE: 24-NOV-99  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 241 amino acids  
TYPE: amino acid  
MOLECULE TYPE: PROTEIN  
SEQUENCE DESCRIPTION: SEQ ID NO: 25:

US-09-424-620B-25

Query Match 98.1%, Score 255, DB 4, Length 241;  
Best Local Similarity 100.0%, Pred. No. 3.3e-29;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNP 49  
|||||  
DB 51 FPTPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNP 98

RESULT: 11  
US-09-280-030-66  
Sequence 66, Application US/092800030A  
Patent No. 653595

GENERAL INFORMATION:  
APPLICANT: Sato, Seiji  
APPLICANT: Higashikuni, Naohiko  
APPLICANT: Kado, Toshiyuki  
APPLICANT: Kondo, Masaaki

TITLE OF INVENTION: DNAS ENCODING NEW FUSION PROTEINS AND PROCESSES FOR THE  
TITLE OF INVENTION: DNAS  
FILE REFERENCE: 382.1026  
CURRENT FILING DATE: 1999-03-26  
EARLIER APPLICATION NUMBER: JP10-87339/1998  
EARLIER FILING DATE: 1998-03-31  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 66  
LENGTH: 245  
TYPE: PRT

ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Designated is  
OTHER INFORMATION: an amino acid sequence of MWpmp20-TEV-G-GR

US-09-280-030-66

Query Match 98.1%, Score 255, DB 4, Length 245;  
Best Local Similarity 100.0%, Pred. No. 3.3e-29;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNP 49  
|||||  
DB 55 FPTPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNP 102

RESULT 12  
US-08-784-582-71  
Sequence 71, Application US/08784582  
Patent No. 6110707

GENERAL INFORMATION:  
APPLICANT: Newgard, Christopher B.  
APPLICANT: Halban, Philippe A.  
APPLICANT: NO. 6110707minqton, Karl D.  
APPLICANT: Clark, Samuel A.  
APPLICANT: Thigpen, Anice E.  
APPLICANT: Osuade, Christian  
APPLICANT: Kruse, Fred  
APPLICANT: McGarity, Dennis

TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM  
TITLE OF INVENTION: SECRETORY CELL LINES  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/784,582  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/028,427  
FILING DATE: 15-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/589,028  
FILING DATE: 19-JAN-1996

ATTORNEY/AGENT INFORMATION:  
NAME: Highlander, Steven L.  
REGISTRATION NUMBER: 37,642  
REFERENCE/DOCKET NUMBER: UTS0:514  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 512/418-3000  
TELEFAX: 512/474-7577

INFORMATION FOR SEQ ID NO: 71:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 274 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear

US-08-784-582-71

Query Match 98.1%, Score 255, DB 3, Length 274;  
Best Local Similarity 100.0%, Pred. No. 3.8e-29;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNP 49  
|||||  
DB 27 FPTPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNP 74

RESULT 13  
US-08-784-582-73  
Sequence 73, Application US/08784582  
Patent No. 6110707

GENERAL INFORMATION:

APPLICANT: Newgard, Christopher B.  
APPLICANT: Halban, Philippe A.  
APPLICANT: No. 6110707mington, Karl D.  
APPLICANT: Clark, Samuel A.  
APPLICANT: Thigpen, Anice E.  
APPLICANT: Quade, Christian  
APPLICANT: Kruse, Fred  
APPLICANT: McGarry, Dennis  
TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM  
TITLE OF INVENTION: SECRETORY CELL LINES  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P. O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPILED: IBM PC compatible  
OPERATING SYSTEM: PC DOS/MS DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/784,562  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/028,427  
FILING DATE: 15-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/589,028  
FILING DATE: 19-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Highlander, Steven L.  
REGISTRATION NUMBER: 37,642  
REFERENCE/DOCKET NUMBER: UTSD:514  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 512/418-3000  
TELEFAX: 512/474-7577  
INFORMATION FOR SEQ ID NO: 73:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 360 amino acids  
TYPE: amino acid  
STRANDEDNESS: linear  
TOPOLOGY:  
US-08-784-562-73

Query Match 98.1% Score 255; DB 3; Length 360;  
Best Local Similarity 100.0%; Pred. No. 5.4e-24;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFEFFAYIPKEOKYSFLONP 49  
|||||  
DB 27 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFEFFAYIPKEOKYSFLONP 74

## RESULT 14

US-09-465-461-1  
Sequence 1, Application US/09465461  
Patent No. 6148444  
GENERAL INFORMATION:

APPLICANT: CHAPPEL, Scott  
TITLE OF INVENTION: Human Growth Hormone to stimulate hematopoiesis and immune reconst  
TITLE OF INVENTION: after hematopoietic stem cell transplantation in humans  
FILE REFERENCE: CHAPPEL-6.1  
CURRENT APPLICATION NUMBER: US/09/465,461  
PRIOR FILING DATE: 1999-12-17  
PRIOR FILING DATE: 1998-12-17  
NUMBER OF SEQ ID NOS: 1  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1

LENGTH: 191  
TYPE: PRO  
ORGANISM: homo sapiens  
US-09-465-461-1

Query Match 95.8% Score 249; DB 4; Length 191;  
Best Local Similarity 97.9%; Pred. No. 1.8e-28;  
Matches 47; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFEFFAYIPKEOKYSFLONP 49  
|||||  
DB 1 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFEFFAYIPKEOKYSFLONP 48

## RESULT 15

US-08-187-756C-4  
Sequence 4, Application US/08187756C  
Patent No. 5597709

GENERAL INFORMATION:  
APPLICANT: POSEN, Eli A.  
TITLE OF INVENTION: Human Growth Hormone  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, HAIN, GILFILLAN,  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPILED: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/187,756C  
FILING DATE: January 27, 1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: FERRARO, GREGORY D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-55  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-954-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 217 AMINO ACIDS  
TYPE: AMINO ACID  
STRANDEDNESS:  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-08-187-756C-4

Query Match 95.8% Score 249; DB 1; Length 217;  
Best Local Similarity 97.9%; Pred. No. 2.1e-24;  
Matches 47; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFEFFAYIPKEOKYSFLONP 49  
|||||  
DB 27 FPTIPLSRFDNAMLRAHRLHQLAFDTYQFEFFAYIPKEOKYSFLONP 74

Search completed: September 15, 2003, 12:05:30  
Job time : 12.2401 secs

GenCore version 1.1.6  
Copyright (c) 1993 - 2003 Empugen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 12:03:35 ; Search time 20.5484 Seconds  
(without alignments)  
347.945 Million cell updates/sec

Title: US-09-423-100-1

Perfect score: 260

Sequence: 1 MF TIPLSRFDNMLRAHR.....OFFFEATIPKQYSFLQNP 49

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 641936 seqs, 14591426 residues

Total number of hits satisfying chosen parameters 641936

Minimum LR seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARY

Result No.	Score	Query Match	Length	DB ID	Description
1	260	100.0	49	14	US-10-054-873-1 Sequence 1, Appli
2	260	100.0	92	14	US-10-054-873-2 Sequence 2, Appli
3	260	100.0	107	14	US-10-054-873-6 Sequence 6, Appli
4	260	100.0	150	14	US-10-054-873-7 Sequence 7, Appli
5	255	98.1	191	11	US-09-984-010-23 Sequence 23, Appli
6	255	98.1	191	12	US-10-153-257-1 Sequence 1, Appli
7	255	98.1	191	12	US-10-400-77-1 Sequence 1, Appli
8	255	98.1	191	12	US-10-400-4-1 Sequence 1, Appli
9	255	98.1	214	12	US-10-153-7-6 Sequence 6, Appli
10	255	98.1	217	9	US-09-929-61-9 Sequence 9, Appli
11	255	98.1	245	9	US-09-280-01-66 Sequence 66, Appli
12	249	95.8	217	9	US-09-853-648-2 Sequence 2, Appli
13	249	95.8	217	12	US-09-969-48C-4 Sequence 4, Appli
14	248	95.4	217	10	US-09-804-409A-16 Sequence 16, Appli
15	242	93.1	217	9	US-09-853-688-4 Sequence 4, Appli

16	239	91.9	191	12	US-09-824-200-12	Sequence 12, Appli
17	199	76.5	217	9	US-09-850-887-3	Sequence 3, Appli
18	197	75.8	163	15	US-10-043-487-150	Sequence 350, App
19	197	75.8	191	12	US-10-153-207-2	Sequence 2, Appli
20	197	75.8	229	15	US-10-103-313-411	Sequence 411, App
21	188	72.3	246	15	US-10-188-246-18	Sequence 18, Appli
22	159.5	61.3	190	12	US-10-153-207-3	Sequence 3, Appli
23	142	54.6	54	9	US-09-876-478-14	Sequence 14, Appli
24	114	43.8	54	9	US-09-876-478-3	Sequence 2, Appli
25	114	43.8	54	9	US-09-876-478-7	Sequence 7, Appli
26	106	40.8	28	9	US-09-876-478-4	Sequence 4, Appli
27	106	40.8	28	9	US-09-876-478-6	Sequence 6, Appli
28	106	40.8	54	9	US-09-876-478-15	Sequence 15, Appli
29	88	33.6	41	15	US-10-191-879-22	Sequence 22, Appli
30	88	33.8	188	15	US-10-191-879-19	Sequence 19, Appli
31	88	33.8	210	15	US-10-191-879-10	Sequence 10, Appli
32	87	33.5	366	10	US-09-887-569A-2	Sequence 2, Appli
33	87	33.5	388	12	US-10-322-746-4	Sequence 4, Appli
34	84	31.9	56	9	US-09-876-478-15	Sequence 15, Appli
35	63	30.5	198	15	US-10-143-293-32	Sequence 32, Appli
36	60	30.6	199	15	US-10-143-293-34	Sequence 33, Appli
37	73	28.1	46	15	US-10-100-679-61	Sequence 61, Appli
38	73	28.1	46	15	US-10-100-679-62	Sequence 62, Appli
39	73	28.1	46	15	US-10-100-679-63	Sequence 63, Appli
40	72.5	27.9	125	14	US-10-036-869-25	Sequence 25, Appli
41	72.5	27.9	253	14	US-10-036-869-27	Sequence 27, Appli
42	71	27.3	197	15	US-10-140-293-13	Sequence 13, Appli
43	70	26.9	199	15	US-10-140-293-22	Sequence 22, Appli
44	70	26.9	199	15	US-10-140-293-23	Sequence 23, Appli
45	70	26.9	199	15	US-10-140-293-25	Sequence 25, Appli

#### ALIGNMENTS

RESULT 1  
US-10-054-873-1  
: Sequence 1, Application US/10054873  
: Publication No. US20020164712A1  
: GENERAL INFORMATION:  
: APPLICANT: Gan, Zhong Ru  
: TITLE OF INVENTION: Intramolecular Protein Containing an  
: NUMBER OF SEQUENCES: 7  
: CORRESPONDENCE ADDRESS:  
: ADDRESS: Townsend and Townsend and Crew LLP  
: STREET: Two Embarcadero Center, Eighth Floor  
: CITY: San Francisco  
: STATE: California  
: COUNTRY: USA  
: ZIP: 94111-3834  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/10/054,873  
: FILING DATE: 22 Jan-2002  
: CLASSIFICATION: <unknown>  
: PRIORITY APPLICATION DATA:  
: APPLICATION NUMBER: WO 98/00052  
: FILING DATE: 31-MAR-1998  
: APPLICATION NUMBER: US 09/423,100  
: FILING DATE: 11-DEC-2000  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Mycroft, Frank J  
: REGISTRATION NUMBER: 46,946  
: REFERENCE/DOCKET NUMBER: 020167-000100US  
: INFORMATION FOR SEQ ID NO: 1:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 49 amino acids  
: TYPE: amino acid

```

; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-054-873-1

Query Match 100.0%; Score 260; DB 14; Length 49;
Best Local Similarity 100.0%; Pred. No. 1 2e-28;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFPTIPLSRFDNAMLRAHRLHQLAFDTYQFEFFAYIPKECKYSFLQNP 49
|||||
DB 1 MFPTIPLSRFDNAMLRAHRLHQLAFDTYQFEFFAYIPKECKYSFLQNP 49
|||||

RESULT 2
US-10-054-873-2
; Sequence 2, Application US/10054873
; Publication No. US20020164712A1
; GENERAL INFORMATION:
; APPLICANT: Gan, Zheng Ku
; TITLE OF INVENTION: Chimeric Protein Containing an
; Intramolecular Chaperone-Like Sequence
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/054-873
; FILING DATE: 22-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/CN98 0952
; FILING DATE: 31-MAR-1998
; APPLICATION NUMBER: US 09/423,100
; FILING DATE: 11-DEC-2000
; ATTORNEY/AGENT INFORMATION:
; NAME: Mycroft, Frank J
; REGISTRATION NUMBER: 46,946
; REFERENCE/DOCKET NUMBER: 620167 - 101300S
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 92 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-054-873-2

Query Match 100.0%; Score 260; DB 14; Length 92;
Best Local Similarity 100.0%; Pred. No. 1 2e-27;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFPTIPLSRFDNAMLRAHRLHQLAFDTYQFEFFAYIPKECKYSFLQNP 49
|||||
DB 1 MFPTIPLSRFDNAMLRAHRLHQLAFDTYQFEFFAYIPKECKYSFLQNP 49
|||||

RESULT 3
US-10-054-873-6
; Sequence 6, Application US/10054873
; Publication No. US20020164712A1
; GENERAL INFORMATION:

```

```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/CN98/00052
: FILING DATE: 31-MAR-1998
: APPLICATION NUMBER: US 09/423,100
: FILING DATE: 11-DEC-2000
: ATTORNEY/AGENT INFORMATION:
: NAME: Mycroft, Frank J
: REGISTRATION NUMBER: 45,946
: REFERENCE/DOCKET NUMBER: 020167-000130US
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 150 amino acids
: TYPE: amino acid
: STRANDEDNESS: <unknown>
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-10-054-873-7

Query Match 100.0%; Score 260; DB 14; Length 150;
Best Local Similarity 100.0%; Pred. No. 1.3e-26;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEPTPLSRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLQNP 49
|||||
Db 1 MEPTPLSRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLQNP 49
|||||

RESULT 5
US-09-984-010-23
: Sequence 23, Application US/09984010
: Publication No. US20030104578A1
: GENERAL INFORMATION:
: APPLICANT: Ballance, David James
: TITLE OF INVENTION: RECOMBINANT FUSION PROTEINS TO GROWTH HORMONE
AND SERUM ALBUMIN
: NUMBER OF SEQUENCES: 25
: CORRESPONDENCE ADDRESS:
: ADDRESS: FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNEP, LLP
: STREET: 1300 I Street, NW
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DCS
: SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/984,010
: FILING DATE: 21-May-2002
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 09/091,873
: FILING DATE: 25-JUN-1998
: APPLICATION NUMBER: PCT/CN96/03164
: FILING DATE: 19-DEC-1996
: INFORMATION FOR SEQ ID NO: 23:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 191 amino acids
: TYPE: amino acid
: STRANDEDNESS: <unknown>
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-984-010-23

Query Match 98.1%; Score 255; DB 11; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.3e-26;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/CN98/00052
: FILING DATE: 31-MAR-1998
: APPLICATION NUMBER: US 09/423,100
: FILING DATE: 11-DEC-2000
: ATTORNEY/AGENT INFORMATION:
: NAME: Mycroft, Frank J
: REGISTRATION NUMBER: 45,946
: REFERENCE/DOCKET NUMBER: 020167-000130US
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 150 amino acids
: TYPE: amino acid
: STRANDEDNESS: <unknown>
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-10-054-873-7

Query Match 100.0%; Score 260; DB 14; Length 150;
Best Local Similarity 100.0%; Pred. No. 1.3e-26;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEPTPLSRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLQNP 49
|||||
Db 1 MEPTPLSRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLQNP 49
|||||

RESULT 6
US-10-153-207-1
: Sequence 1, Application US/10153207
: Publication No. US20030153003A1
: GENERAL INFORMATION:
: APPLICANT: James A. Wells
: APPLICANT: Brian C. Cunningham
: TITLE OF INVENTION: GROWTH HORMONE VARIANTS
: FILE REFERENCE: 669,12-US-C7
: CURRENT APPLICATION NUMBER: US/10/153,207
: PRIOR FILING DATE: 2002-05-22
: PRIOR APPLICATION NUMBER: 08/479,884
: PRIOR FILING DATE: 1995-06-07
: PRIOR APPLICATION NUMBER: 08/196,723
: PRIOR FILING DATE: 1994-02-02
: PRIOR APPLICATION NUMBER: 07/969,227
: PRIOR FILING DATE: 1992-10-13
: PRIOR APPLICATION NUMBER: 07/875,204
: PRIOR FILING DATE: 1992-04-27
: PRIOR APPLICATION NUMBER: 07/428,066
: PRIOR FILING DATE: 1989-10-26
: PRIOR APPLICATION NUMBER: 07/264,611
: PRIOR FILING DATE: 1988-10-28
: NUMBER OF SEQ ID NOS: 20
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 191
: TYPE: PRT
: ORGANISM: Homo Sapiens
US-10-153-207-1

Query Match 98.1%; Score 255; DB 12; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.3e-26;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PPTPLSRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLQNP 49
|||||
Db 1 PPTPLSRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLQNP 49
|||||

RESULT 7
US-10-400-377-1
: Sequence 1, Application US/10400377
: Publication No. US20030162949A1
: GENERAL INFORMATION:
: APPLICANT: Cox III, George N
: APPLICANT: Bolder Biotechnology, Inc.
: TITLE OF INVENTION: Derivatives of Growth Hormone and Related Proteins
: FILE REFERENCE: 4152-1-PUS
: CURRENT APPLICATION NUMBER: US/10/400,377
: CURRENT FILING DATE: 2003-03-26
: PRIOR APPLICATION NUMBER: US/09/462,941
: PRIOR FILING DATE: 2000-01-14
: PRIOR APPLICATION NUMBER: 60/052,516
: PRIOR FILING DATE: 1997-07-14
: NUMBER OF SEQ ID NOS: 41
: SOFTWARE: Patent In Ver. 2.0
: SEQ ID NO 1
: LENGTH: 191
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-400-377-1

Query Match 98.1%; Score 255; DB 12; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.3e-26;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PPTPLSRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLQNP 49
|||||
Db 1 PPTPLSRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLQNP 49
|||||
```

Db 1 FTPIPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNP 48  
|||||  
RESULT 8  
US-10-400-708-1  
; Sequence 1, Application US/10400708  
; Publication No. US2003016863A1  
; GENERAL INFORMATION:  
; APPLICANT: Cox III, George N  
; TITLE OF INVENTION: Derivatives of Growth Hormone and Related Proteins  
; FILE REFERENCE: 4152-1-PUS  
; CURRENT APPLICATION NUMBER: US/10/400,708  
; CURRENT FILING DATE: 2003-03-26  
; PRIOR APPLICATION NUMBER: US/09/462,941  
; PRIOR FILING DATE: 2000-01-14  
; PRIOR APPLICATION NUMBER: 60/052,516  
; PRIOR FILING DATE: 1997-07-14  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-400-708-1

Query Match 98.1%; Score 255; DB 12; Length 191;  
Best Local Similarity 100.0%; Pred. No. 1.5e-26;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 FTPIPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNP 49  
Db 1 FTPIPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNP 48  
|||||

RESULT 9  
US-10-151-207-6  
; Sequence 6, Application US/10153207  
; Publication No. US20030153003A1  
; GENERAL INFORMATION:  
; APPLICANT: James A. Wells  
; APPLICANT: Brian C. Cunningham  
; TITLE OF INVENTION: GROWTH HORMONE VARIANT  
; FILE REFERENCE: 669-12 US-C7  
; CURRENT APPLICATION NUMBER: US/10/153,207  
; CURRENT FILING DATE: 2002-05-22  
; PRIOR APPLICATION NUMBER: 08/479,884  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 08/190,723  
; PRIOR FILING DATE: 1994-02-02  
; PRIOR APPLICATION NUMBER: 07/960,227  
; PRIOR FILING DATE: 1992-10-11  
; PRIOR APPLICATION NUMBER: 07/875,204  
; PRIOR FILING DATE: 1992-04-27  
; PRIOR APPLICATION NUMBER: 07/428,056  
; PRIOR FILING DATE: 1989-10-26  
; PRIOR APPLICATION NUMBER: 07/264,611  
; PRIOR FILING DATE: 1988-10-28  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 214  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
US-10-151-207-6

Query Match 98.1%; Score 255; DB 12; Length 214;  
Best Local Similarity 100.0%; Pred. No. 1.5e-26;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 FTPIPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNP 49  
|||||

Db 24 FTPIPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNP 71  
|||||  
RESULT 10  
US-09-929-918-9  
; Sequence 9, Application US/09929918  
; Patent No. US20020090678A1  
; GENERAL INFORMATION:  
; APPLICANT: Kordyum, Vitaliy A.  
; APPLICANT: Chervykh, Svitlana I.  
; APPLICANT: Slavchenko, Iryna Yu.  
; APPLICANT: Vozianov, Oleksandr  
; TITLE OF INVENTION: PHAGE-DEPENDENT SUPER PRODUCTION OF  
; FILE REFERENCE: BIOLOGICALLY ACTIVE PROTEIN AND PEPTIDES  
; CURRENT APPLICATION NUMBER: US/09/929,918  
; CURRENT FILING DATE: 2001-08-15  
; PRIOR APPLICATION NUMBER: 09/314,288  
; PRIOR FILING DATE: 1999-05-25  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 217  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-929-918-9

Query Match 98.1%; Score 255; DB 9; Length 217;  
Best Local Similarity 100.0%; Pred. No. 1.5e-26;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 FTPIPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNP 49  
Db 27 FTPIPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNP 74  
|||||

RESULT 11  
US-09-280-030-66  
; Sequence 66, Application US/092803030A  
; Patent No. US20010021515A1  
; GENERAL INFORMATION:  
; APPLICANT: Sato, Seiji  
; APPLICANT: Higashikuni, Naohiko  
; APPLICANT: Kudo, Toshiyuki  
; APPLICANT: Kondo, Masaaki  
; TITLE OF INVENTION: DNAS ENCODING NEW FUSION PROTEINS AND PROCESSES FOR THE  
; FILE REFERENCE: DNAS  
; CURRENT APPLICATION NUMBER: 382,1026  
; CURRENT FILING DATE: 1999-03-26  
; EARLIER APPLICATION NUMBER: JP10-87345/1998  
; EARLIER FILING DATE: 1998-03-31  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 66  
; LENGTH: 245  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Designated as  
; OTHER INFORMATION: an amino acid sequence of MWpsp-MWp20-TEV-G-SH  
US-09-280-030-66

Query Match 98.1%; Score 255; DB 9; Length 245;  
Best Local Similarity 100.0%; Pred. No. 1.7e-26;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 FTPIPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNP 49  
Db 55 FTPIPLSRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEOKYSFLQNP 102  
|||||



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:56:45 ; Search time 8.07885 Seconds  
(without alignments)  
583.284 Million cell updates/sec

Title: US-09-423-100-1

Perfect score: 260

Sequence: 1 MFPTIPLSRLFDNAHLRAH.....QEFEEVYIPKQKYSFTLQNP 49

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 46168642 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Query Match	Length	ID	Description
1	255	98.1	217	1	STHO
2	255	98.1	217	2	somatotropin 1 pre
3	228	87.7	217	1	STHUV
4	228	87.7	256	1	somatotropin 2 pre
5	213	81.9	212	2	somatotropin 2 pre
6	213	81.9	212	2	chorionic somatoma
7	205	78.8	217	2	chorionic somatoma
8	205	78.8	217	2	somatotropin - the
9	197	75.8	215	2	chorionic somatoma
10	197	75.8	215	2	chorionamniotropon
11	197	75.8	217	1	LCHUC
12	161.5	62.1	216	1	STWS
13	160.5	61.7	190	2	PN0140
14	159.5	61.3	190	1	STHO
15	159.5	61.3	190	2	JK0219
16	159.5	61.3	190	2	J50429
17	159.5	61.3	216	1	STRT
18	159.5	61.3	216	1	STPG
19	159.5	61.3	216	2	146145
20	159.5	61.3	216	2	S49483
21	159.5	61.3	216	2	B49159
22	159.5	61.3	216	2	JC4632
23	156.5	60.2	216	2	A37782
24	155.5	59.8	190	1	A61584
25	150	57.7	216	2	JC1514
26	148	56.9	191	2	A60625
27	146	56.2	163	2	JN0387
28	144	55.4	190	2	S21750
29	144	55.4	216	2	A60509

30	142.5	54.8	217	1	STBO
31	142.5	54.8	217	1	STSH
32	142.5	54.8	217	1	STGT
33	142.5	54.8	217	2	S32682
34	140	53.8	216	2	S04929
35	132	50.8	190	2	A56816
36	132	50.8	215	2	I51188
37	128	49.2	195	2	I51250
38	128	49.2	215	2	J50037
39	122	46.4	199	2	B32435
40	116	44.6	183	2	A60623
41	98.5	37.9	87	4	I67761
42	97	37.3	200	2	I51114
43	87	33.5	210	2	S69263
44	87	33.5	210	2	S69262
45	87	33.5	210	2	S02764

# ALIGNMENTS

## RESULT 1

STHO

somatotropin 1 precursor [validated] - human  
N:Alternate names: growth hormone 1; hGH-N; pituitary somatotropin  
N:Contains: growth hormone 5K peptide; somatotropin 1, long form; somatotropin 1, st  
C:Species: Homo sapiens (man)  
C:Date: 24-Apr-1984 #sequence-revision 10-Feb 1995 #text-change 08-Dec-2000  
C:Accession: A93731; A32435; A93694; A94247; A90051; A93397; A93778; A91764; A90217;  
R:DeMoto, F.M.; Moore, D.D.; Goodman, H.M.  
Nucleic Acids Res. 9, 1719-1730, 1981  
A:Title: Human growth hormone DNA sequence and mRNA structure: possible alternative  
A:Reference number: A93731; MUID:82014939; PMID:6249091  
A:Accession: A93731  
A:Molecule type: DNA  
A:Residues: 1-217 <DEN>  
A:Cross-references: GB:V00520  
R:Chen, E.Y.; Liao, Y.C.; Smith, D.H.; Barrera-Saldana, H.A.; Gelinas, R.E.; Seebur-  
Genomics 4, 479-497, 1989  
A:Title: The human growth hormone locus: nucleotide sequence, biology, and evolution  
A:Reference number: A32435; MUID:83307277; PMID:2744760  
A:Accession: A32435  
A:Molecule type: DNA  
A:Residues: 1-217 <CH>  
A:Cross-references: GB:J03071; NID:q183148; PIDN:AAA52549.1; PID:q183149  
R:Roskam, W.; Rougeon, F.  
Nucleic Acids Res. 7, 305-320, 1979  
A:Title: Molecular cloning and nucleotide sequence of the human growth hormone struc-  
A:Reference number: A93694; MUID:80034477; PMID:386281  
A:Accession: A93694  
A:Molecule type: mRNA  
A:Residues: 1-217 <RGS>  
A:Cross-references: GB:V00519  
R:Martial, J.A.; Halliwell, R.A.; Baxter, J.D.; Goodman, H.M.  
Science 205, 602-607, 1979  
A:Title: Human growth hormone: complementary DNA cloning and expression in bacteria  
A:Reference number: A94247; MUID:79203293; PMID:377496  
A:Accession: A94247  
A:Molecule type: mRNA  
A:Residues: 1-217 <MAR>  
R:Li, C.H.; Dixon, J.S.; Liu, W.K.  
Arch. Biochem. Biophys. 133, 70-91, 1969  
A:Title: Human pituitary growth hormone. XIX. The primary structure of the hormone.  
A:Reference number: A90048; MUID:69289202; PMID:5810834  
A:Contents: annotation  
R:Li, C.H.; Dixon, J.S.  
Arch. Biochem. Biophys. 146, 233-236, 1971  
A:Title: Human pituitary growth hormone. XXXI. The primary structure of the hormone.  
A:Reference number: A90051; MUID:72143935; PMID:5144027  
A:Accession: A90051  
A:Molecule type: protein



A:Residues: 27-94;96-217 <LIC>  
R:Niall, H.D.  
Nature New Biol. 230, 90-91, 1971  
A:Title: Revised primary structure for human growth hormone.  
A:Reference number: A93397; MUID:71139765; PMID:5279046  
A:Accession: A93397  
A:Molecule type: protein  
A:Residues: 27-51 <NIA>  
R:Niall, H.D.; Hogan, M.L.; Sauer, R.; Rosenblum, I.V.; Greenwood, F.C.  
Proc. Natl. Acad. Sci. U.S.A. 68, 866-869, 1971  
A:Title: Sequences of pituitary and placental lactogenic and growth hormones: evolution  
A:Reference number: A93778; MUID:71153968; PMID:5279528  
A:Accession: A93778  
A:Molecule type: protein  
A:Residues: 119-120;157-159 <NIZ>  
R:Niall, H.D.  
in Prolactin and Carcinogenesis, Proc. Fourth Tenovus Workshop Prolactin, Griffiths, K.,  
A:Title: The Chemistry of the human lactogenic hormones.  
A:Reference number: A94427  
R:Contents: annotation; somatotropin revision  
R:Bewley, T.A.; Jirik, J.S.; Liu, C.H.  
Int. J. Pept. Protein Res. 4, 281-287, 1972  
A:Title: Sequence comparison of human pituitary growth hormone, human chorionic somatom-  
A:Reference number: A91764; MUID:7302628; PMID:477544  
A:Accession: A91764  
A:Molecule type: protein  
A:Residues: 27-217 <BEW>  
R:Lewis, U.J.; Bonawald, L.F.; Lewis, L.J.  
Biochem. Biophys. Res. Commun. 92, 511-516, 1980  
A:Title: The 20,000-dalton variant of human growth hormone: location of the amino acid  
A:Reference number: A90217; MUID:80130196; PMID:7356479  
A:Contents: somatotropin, 20K short variant  
A:Accession: A90217  
A:Molecule type: protein  
A:Residues: 45-57;74-80 <LEW>  
R:Chapman, G.E.; Rogers, K.M.; Brittain, T.; Bradshaw, R.A.; Bates, G.L.; Turner, C.; Ca-  
J. Biol. Chem. 256, 2395-2401, 1981  
A:Title: The 20,000 molecular weight variant of human growth hormone. Preparation and so-  
A:Reference number: A92311; MUID:81117361; PMID:7462247  
A:Contents: somatotropin, 20K short variant  
A:Accession: A92311  
A:Molecule type: protein  
A:Residues: 27-57;73-79 <CHA>  
R:Singh, R.N.P.; Seavey, B.K.; Lewis, L.J.; Lewis, H.J.  
J. Protein Chem. 2, 425-436, 1983  
A:Title: Human growth hormone peptide 1-43: isolation from pituitary glands.  
A:Reference number: A61466  
A:Accession: A61466  
A:Molecule type: protein  
A:Residues: 27-69 <SIN>  
R:Robson, V.M.J.; Rae, I.D.; NG, F.  
Biol. Chem. Hoppe-Seyler 371, 423-431, 1990  
A:Title: Identification of the aspartamide structure in a previously-reported peptide.  
A:Reference number: S09685; MUID:90334745; PMID:2378679  
A:Accession: S09685  
A:Molecule type: protein  
A:Residues: 27-14;17,16-47 <ROB>  
R:De Vos, A.M.; Uitsch, M.; Kossiakoff, A.A.  
Science 255, 306-312, 1992  
A:Title: Human growth hormone and extracellular domain of its receptor: crystal structu-  
A:Reference number: A41728; MUID:92196577; PMID:1549776  
A:Contents: annotation; X-ray crystallography, 2.8 angstroms  
A:Note: The structure of the complex with growth hormone receptor is described  
R:Gray, G.L.; Baldrige, J.S.; McKeown, K.S.; Heyneker, H.L.; Chang, C.N.  
Gene 19, 247-254, 1985  
A:Title: Periplasmic production of correctly processed human growth hormone in Escherich-  
A:Reference number: I41126; MUID:86137393; PMID:3912261  
A:Accession: I84549  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-26 <RES>  
A:Cross-references: GB:M14398; NID:g183158; PIDN:AAA52554.1; PMID:g183159

C:Comment: The gene for this hormone is transcribed only in somatotrophic cells of th  
C:Comment: About 90% of somatotropin is the 22K long form.  
C:Genetics:  
A:Gene: GDB:GH1  
A:Cross-references: GDB:119982; OMIM:139250  
A:Map position: 17q23.1-17q23.3  
A:Introns: 4/1; 57/3; 97/3; 152/3  
C:Superfamily: prolactin  
C:Keywords: alternative splicing; hormone; pituitary  
F:1-26/Domain: signal sequence #status predicted <SIG>  
F:27-217/Product: somatotropin 1, long form #status experimental <SIG>  
F:27-57/Product: growth hormone 5K peptide #status experimental <SKP>  
F:27-57;73-217/Product: somatotropin 1, short form #status experimental <SOS>  
F:79-191;208-215/Disulfide bonds: #status experimental

Query Match 98.1%; Score 255; DB 1; Length 217;  
Best Local Similarity 100.0%; Pred. No. 2e-24;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTIFSLFNNAMLRRAHKLHQLAFDTYQFFFEAYIPKEQKYSFLQNP 49  
DB 27 FPTIFSLFNNAMLRRAHKLHQLAFDTYQFFFEAYIPKEQKYSFLQNP 74

RESULT 2  
167410  
somatotropin - rhesus macaque  
N:Alternate names: growth hormone  
C:Species: Macaca mulatta (rhesus macaque)  
C>Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 16-Jul-1999  
C:Accession: I67410; A05094  
R:Golos, T.G.; Burningham, M.; Fisher, J.M.; Fowler, P.D.  
Endocrinology 133, 1744-1752, 1993  
A:Title: Cloning of four growth hormone/chorionic somatomammotropin-related complen-  
A:Reference number: I53267; MUID:94008724; PMID:8414617  
A:Accession: I67410  
A>Status: translated from GB/EMBL/DDHJ  
A:Molecule type: mRNA  
A:Residues: 1-217 <RES>  
A:Cross-references: GB:IJ6556; NID:g293114; PIDN:AAA18842.1; PID:g293115  
R:Li, C.H.; Chung, D.; Lahm, H.W.; Stein, S.  
Arch. Biochem. Biophys. 245, 287-291, 1986  
A:Title: The primary structure of monkey pituitary growth hormone.  
A:Reference number: A05094; MUID:86129460; PMID:3080959  
A:Accession: A05094  
A:Molecule type: protein  
A:Residues: 27-99; Q'101-178; D'180-217 <LIC>  
A:Note: The monkey species is not identified in the reference  
R:Raben, M.S.  
Science 125, 883-884, 1957  
A:Title: Preparation of growth hormone from pituitaries of man and monkey.  
A:Reference number: A44774  
A:Contents: annotation; identification of source organism  
C:Superfamily: prolactin

Query Match 98.1%; Score 255; DB 2; Length 217;  
Best Local Similarity 100.0%; Pred. No. 2e-24;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTIFSLFNNAMLRRAHKLHQLAFDTYQFFFEAYIPKEQKYSFLQNP 49  
DB 27 FPTIFSLFNNAMLRRAHKLHQLAFDTYQFFFEAYIPKEQKYSFLQNP 74

RESULT 3  
STHUV  
somatotropin 2 precursor - human  
N:Alternate names: growth hormone 2; growth hormone variant; hGH-V; placental somato-  
N:Contents: somatotropin 2, long splice form; somatotropin 2, short splice form  
C:Species: Homo sapiens (man)  
C>Date: 17-Dec-1982 #sequence\_revision 10-Feb-1995 #text\_change 21-Jul-2000  
C:Accession: D32435; R28072; A01511; I52104; A6071;  
R:Chen, F.Y.; Liao, Y.C.; Smith, D.H.; Barrera-Saldana, H.A.; Gelinas, R.E.; Seeburg

A:Accession: A28072  
A:Molecule type: mRNA  
A:Residues: 1-256 <COO>  
A:Note: an alternative splice junction for intron 4 is used  
C:Genetics:  
A:Gene: GDB:GH2  
A:CROSS-references: GDB:I19983; OMIM:139240  
A:Map position: 17q22-17q24  
A:Introns: 4/1; 57/3; 97/3; 152/3  
C:Superfamily: prolactin  
C:Keywords: alternative splicing; hormone; placenta  
F:1-26/Domain: signal sequence \*status predicted <SIG>  
F:27-256/Product: somatotropin 2 splice form 2 \*status predicted <MAT>

Query Match 87.7% Score 228; DB 1: Length 256;  
Best local similarity 91.7%; Pred. No. 5,6e-21;  
Matches 44; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

CY 2 PTIPSLRFDNAMIQAHRHLQAFDTYOFEFFAYIPKEKYSFLQNPF 49  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 27 PVPPLSRFLDAMIQAHRLHQLAFDTYOFEFFAYIPKEKYSFLQNP 74

RESULT 5  
I67408  
Chorionic somatomammotropin-2 - rhesus macaque (:raymont)  
C:Species: Macaca mulatta (rhesus macaque)  
C:Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 16-Jul-1999  
C:Accession: I67408  
R:Colos, T.G.; Durning, M.; Fisher, J.M.; Fowler, P.D.  
Endocrinology 133, 1744-1752, 1993  
A:Title: Cloning of four growth hormone/chorionic somatomammotropin-related compleme  
A:Reference number: I53267; MUID:94008724; PMID:8404617  
A:Accession: I67408  
A>Status: preliminary; translated from GR/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-212 <RES>  
A:CROSS-references: GDB:I16553; NID:g293110; PIDN:AAA18840.1; PID:g293111  
C:Superfamily: prolactin

Query Match 81.9% Score 213; DB 2: Length 212;  
Best local similarity 78.7%; Pred. No. 3,4e-11;  
Matches 37; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

CY 3 PTIPSLRFDNAMIQAHRHLQAFDTYOFEFFAYIPKEKYSFLQNPF 49  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 23 PSVPLSRFLDAMIQAHRLHQLAFDTYOFEFFAYIPKEKHSLSMENP 69

RESULT 6  
I53267  
Chorionic somatomammotropin-1 - rhesus macaque  
C:Species: Macaca mulatta (rhesus macaque)  
C:Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 16-Jul-1999  
C:Accession: I53267  
R:Colos, T.G.; Durning, M.; Fisher, J.M.; Fowler, P.D.  
Endocrinology 133, 1744-1752, 1993  
A:Title: Cloning of four growth hormone/chorionic somatomammotropin-related complem  
A:Reference number: I53267; MUID:94008724; PMID:8404617  
A:Accession: I53267  
A>Status: preliminary; translated from GR/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-217 <RES>  
A:CROSS-references: GDB:I16552; NID:g293108; PIDN:AAA18839.1; PID:g293109  
C:Superfamily: prolactin

Query Match 81.9% Score 213; DB 2: Length 217;  
Best local similarity 78.7%; Pred. No. 3,5e-19;  
Matches 37; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

CY 3 PTIPSLRFDNAMIQAHRHLQAFDTYOFEFFAYIPKEKYSFLQNPF 49  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 28 PSVPLSRFLDAMIQAHRLHQLAFDTYOFEFFAYIPKEKHSLSMENP 74

```

Best Local Similarity 80.0%; Pred. No. 3,40-17;
Matches 36; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 4 TIPLSRFLDNAMLRHRLHQLAFDTYQFEFFAYIPKEQKYSFLON 48
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 29 TVPLSRFLDNAMLRHRLHQLAFDTYQFEFFAYIPKEQKYSFLND 73

RESULT 10
LCHUC
Chorionamniotropin A precursor [validated] - human
N:Alternate names: chorionic somatomammotropin 1; placental lactogen
C:Species: Homo sapiens (man)
C:Date: 23-Oct-1981 #sequence_revision 23-Oct-1981 #text_change 08-Dec-2000
A:Reference number: A93833; A93833; A93192; A90054; A94427; A61283; 155229;
R:Chen, E.Y.; Liao, Y.C.; Smith, D.H.; Barrera-Saldana, H.A.; Gellinas, R.E.; Seeburg,
Genomics 4, 479-497, 1989
A:Title: The human growth hormone locus: nucleotide sequence, biology, and evolution.
A:Reference number: A92445; MUID:8207277; PMID:2744765
A:Accession: A92445
A:Molecule type: DNA
A:Residues: 1-217 <CH>
A:Cross-references: GB:J63071; NID:q183146; PID:AAA52551.1; PID:q183151
R:Goodman, H.M.; DeNoto, F.; Fiddes, J.C.; Hillewell, R.A.; Page, G.S.; Smith, S.; Ti
in Mobilization and Reassembly of Genetic Information, Scott, W.A.; Werner, R.; Josep
A:Reference number: A94422
A:Accession: A94422
A:Molecule type: mRNA
A:Residues: 1-217 <CH>
R:Tanaka, M.; Masuda, N.; Watabiki, M.; Yamakawa, M.; Shimizu, K.; Naqai, J.; Nakashu
Biochem. Int. 16, 287-292, 1988
A:Title: cDNA cloning of human chorionic somatomammotropin-1 cRNA whose transcription:
A:Reference number: 152342; MUID:88209096; PMID:2835050
A:Accession: 152342
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3 <TAN>
R:Sherwood, L.M.; Burstein, Y.; Schechter, I.
Proc. Natl. Acad. Sci. U.S.A. 76, 3819-3823, 1979
A:Title: Primary structure of the NH-2-terminal: extra piece of the precursor to human
A:Reference number: A93833; MUID:80034970; PMID:291043
A:Accession: A93833
A:Molecule type: protein
A:Residues: 1-3-26 <SHE>
A:Experimental source: Placenta
R:Shine, J.; Seeburg, P.H.; Martial, J.A.; Baxter, J.D.; Goodman, H.M.
Nature 270, 494-499, 1977
A:Title: Construction and analysis of recombinant DNA for human chorionic somatamm
A:Reference number: A93192; MUID:78671761; PMID:593368
A:Accession: A93192
A:Molecule type: DNA
A:Residues: 50-217 <SHI>
A:Experimental source: Placenta
R:Li, C.H.; Dixon, J.S.; Chung, D.
Arch. Biochem. Biophys. 155, 95-110, 1973
A:Title: Amino acid sequence of human chorionic somatomammotropin.
A:Reference number: A90054; MUID:74201971; PMID:4712450
A:Accession: A90054
A:Molecule type: protein
A:Residues: 27-217 <LIC>
A:Experimental source: Placenta
R:Niall, H.D.
In Prolactin and Carcinogenesis, Proc. Fourth Tenovus Workshop Prolactin, Griffiths,
A:Title: The chemistry of the human lactogenic hormones.
A:Reference number: A94427
A:Accession: A94427
A:Molecule type: protein
A:Residues: 27-217 <NIA>
R:Nic A Bhaird, N.; Tipton, K.F.
Biochem. Soc. Trans. 19, 20S, 1991
A:Title: Catechol-O-methyltransferase from human placenta: purification and some prop

```

A:Reference number: A61283; MUID:91244006; PMID:2037148  
 A:Accession: A61283  
 A:Molecule type: protein  
 A:Residues: 27-46 <MIC>  
 A:Note: chorionamniotropon apparently copurified with placental catechol-O-methyltransferase  
 R:Sherwood, L.M.; Handwerker, S.; McLaurin, W.D.; Lanner, M.  
 Nature New Biol. 233, 59-61, 1971  
 A:Title: Amino-acid sequence of human placental lactogen.  
 A:Reference number: A93401; MUID:72016313; PMID:5286363  
 A:Accession: A93401  
 R:Sherwood, L.M.; Handwerker, S.; McLaurin, W.D.; Lanner, M.  
 Nature New Biol. 235, 64, 1972  
 A:Reference number: A93405  
 A:Accession: A93405  
 A:Contents: annotation  
 R:Schneider, A.B.; Kowalski, K.; Russell, J.; Sherwood, L.M.  
 J. Biol. Chem. 254, 3782-3787, 1979  
 A:Title: Identification of the interchain disulfide bonds of dimeric human placental lactogen.  
 A:Reference number: A92251; MUID:79173941; PMID:478159  
 A:Accession: A92251  
 R:Selby, M.J.; Barta, A.; Baxter, J.F.; Ellis, J.L.; Farquhar, C. Nils.  
 J. Biol. Chem. 259, 14331-14336, 1984  
 A:Title: Analysis of a major human chorionic somatomammotropin gene. Evidence for two functional alleles.  
 A:Reference number: 155229; MUID:85030426; PMID:6208192  
 A:Accession: 155229  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-217 <RES>  
 A:Cross-references: GB:K02401; NID:g181120 IDN:AAAS2115.1; PID:g181121  
 R:Seeburg, P.H.; Shine, J.; Martial, J.A.; Rich, A.; Goodman, H.  
 Trans. Assoc. Am. Physicians 90, 109-116, 1977  
 A:Title: Nucleotide sequence of a human gene coding for a polypeptide hormone.  
 A:Reference number: 159658; MUID:78.60787; MUID:611657  
 A:Accession: 159658  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 160-217 <RES2>  
 A:Cross-references: GB:M25118; NID:g181124; PIDN:AAAS5721.1; PID:g181125  
 C:Genetics:  
 A:Gene: GBR:CSH1  
 A:Cross-references: GDB:119084; OMIM:150200  
 A:Map position: 17q22-17q24  
 A:Introns: 4/1; 57/3; 97/3; 152/3  
 C:Superfamily: prolactin  
 C:Keywords: hormone; Placenta  
 F:1-26/Domain: signal sequence #status experimental <SIG>  
 F:27-217/Product: chorionamniotropon A #status experimental <MAT>  
 F:79-191/Disulfide bonds: #status experimental  
 F:208-215/Disulfide bonds: (in monomeric form) #status experimental  
 F:208/Disulfide bonds: interchain (to 215 in dimeric form) #status experimental  
 F:215/Disulfide bonds: interchain (to 208 in dimeric form) #status experimental  
 Query Match 75.88; Score 197; DB 1; Length 217;  
 Best Local Similarity 80.0%; Pred. No. 3.5e-17;  
 Matches 36; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
 QY 4 TIPLSLFDNMLRAHRLHQLAFDTYOEFEAYIPKEQKYSFLON 48  
 DB 29 TVPLSKLFDHMLQAHRAHQLAIDTYOEFEAYIPKDKYSFLHD 73  
 RESULT 11  
 E32435  
 Chorionamniotropon B precursor - human  
 N:Alternate names: chorionic somatomammotropin 2  
 C:Species: Homo sapiens (man)  
 C:Date: 24-Dec-1989 #sequence\_revision 29-Dec-1989 #text\_change 16-Jul-1999  
 C:Accession: E32435  
 R:Chen, E.Y.; Liao, Y.C.; Smith, D.H.; Barrera-Saldana, H.A.; Gellinas, R.E.; Seeburg, P.  
 Genomics 4, 479-497, 1989  
 A:Title: The human growth hormone locus: nucleotide sequence, biology, and evolution.  
 A:Reference number: A32435; MUID:89307277; PMID:2744760  
 A:Accession: E32435  
 A:Status: preliminary

A:Molecule type: DNA  
 A:Residues: 1-217 <CHE>  
 A:Cross-references: GB:J03071; NID:g183148; PIDN:AAAS2553.1; PID:g183153  
 C:Genetics:  
 A:Gene: GBR:CSH2  
 A:Cross-references: GDB:119813; OMIM:118820  
 A:Map position: 17q22-17q24  
 C:Superfamily: prolactin  
 Query Match 75.88; Score 197; DB 2; Length 217;  
 Best Local Similarity 80.0%; Pred. No. 3.5e-17;  
 Matches 36; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
 QY 4 TIPLSLFDNMLRAHRLHQLAFDTYOEFEAYIPKEQKYSFLON 48  
 DB 29 TVPLSKLFDHMLQAHRAHQLAIDTYOEFEAYIPKDKYSFLHD 73  
 RESULT 12  
 STMS  
 Somatotropin precursor - mouse  
 N:Alternate names: growth hormone  
 C:Species: Mus musculus (house mouse)  
 C:Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 28-May-1999  
 C:Accession: B23911  
 R:Linzer, D.I.H.; Talamantes, F.  
 J. Biol. Chem. 260, 9574-9579, 1985  
 A:Title: Nucleotide sequence of mouse prolactin and growth hormone mRNAs and expressed cDNAs.  
 A:Reference number: A92548; MUID:85261358; PMID:2991252  
 A:Accession: B23911  
 A:Molecule type: mRNA  
 A:Residues: 1-216 <LJN>  
 A:Cross-references: GB:X02891; GR:K03232; NID:g51067; PIDN:CAA26650.1; PID:g51068  
 C:Superfamily: prolactin  
 C:Keywords: anterior pituitary; growth factor; hormone  
 F:1-26/Domain: signal sequence #status predicted <SIG>  
 F:27-216/Product: somatotropin #status predicted <STN>  
 F:78-189,206-214/Disulfide bonds: #status predicted  
 Query Match 62.1%; Score 161.5; DB 1; Length 216;  
 Best Local Similarity 68.1%; Pred. No. 9.4e-13;  
 Matches 32; Conservative 6; Mismatches 8; Indels 1; Gaps 1;  
 QY 2 FFTIPLSLFDNMLRAHRLHQLAFDTYOEFEAYIPKEQKYSFLON 48  
 DB 27 FPAMP:LSLIFSNAVLRAQLHQLAADTYKEFRAY:PEGORYS-ION 72  
 RESULT 13  
 PN0140  
 Somatotropin - sei whale  
 N:Alternate names: growth hormone  
 C:Species: Balaeoptera borealis (sei whale)  
 C:Date: 07-May-1993 #sequence\_revision 07-May-1993 #text\_change 07-May-1999  
 C:Accession: PN0140  
 R:Yudaev, N.A.; Pankov, Y.A.; Bulatov, A.A.; Osipova, T.A.  
 Biochimia 47, 1059-1069, 1982  
 A:Title: Amino acid sequence of sei whale somatotropin.  
 A:Reference number: PN0140; MUID:83000569; PMID:7115813  
 A:Accession: PN0140  
 A:Molecule type: protein  
 A:Residues: 1-190 <YUD>  
 A:Note: article in Russian with English abstract  
 C:Superfamily: prolactin  
 C:Keywords: growth factor; hormone  
 F:52-163,180-188/Disulfide bonds: #status predicted  
 Query Match 61.7%; Score 160.5; DB 2; Length 190;  
 Best Local Similarity 68.1%; Pred. No. 1.1e-12;  
 Matches 32; Conservative 6; Mismatches 8; Indels 1; Gaps 1;  
 QY 2 FFTIPLSLFDNMLRAHRLHQLAFDTYOEFEAYIPKEQKYSFLON 48  
 DB 27 FPAMP:LSLIFSNAVLRAQLHQLAADTYKEFRAY:PEGORYS-ION 72



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:54:00 ; Search time 4.56631 seconds  
(without alignments)  
504.633 Million cell updates/sec

Title: US-09-423-100-1

Perfect score: 260

Sequence: i MPPTIPLSLFDNAMI,RAHR.....QEFPEAYTPKEQKYSFLQNP 49

Scoring table: HXSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127864 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127864

Minimum hit seq length: 0

Maximum hit seq length: 200000000

Post-process: no; Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Prod No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	255	98.1	217	1 SOMA_HUMAN	P01241 homo sapien
2	255	98.1	217	1 SOMA_MACMU	P33093 macaca mula
3	255	98.1	217	1 SOMA_PANTR	P58756 pan troglod
4	249	95.8	217	1 SOMA_CALJA	O9qmb3 callithrix
5	249	95.8	217	1 SOMA_SALIB	P58343 salmirel boi
6	236	90.8	217	1 SOM2_PANTR	P58757 pan troglod
7	228	87.7	217	1 SOM2_HUMAN	P01242 homo sapien
8	199	76.5	217	1 SOM2_MACMU	P01243 macaca mula
9	197	75.8	217	1 P11_HUMAN	P01243 homo sapien
10	161.5	62.1	216	1 SOMA_MUUSE	P06880 mus musculu
11	160.5	61.7	190	1 SOMA_BALBO	P33092 balaeopter
12	159.5	61.3	190	1 SOMA_LOXAF	P20392 loxodonta a
13	159.5	61.3	190	1 SOMA_VULVU	P10766 vulpes vulp
14	159.5	61.3	216	1 SOMA_CANFA	P33711 canis fami
15	159.5	61.3	216	1 SOMA_FELCA	P46404 felis silve
16	159.5	61.3	216	1 SOMA_HORSE	P01245 equus cabal
17	159.5	61.3	216	1 SOMA_MESAU	P37886 mesocricetu
18	159.5	61.3	216	1 SOMA_PIG	P01248 sus scrofa
19	159.5	61.3	216	1 SOMA_RABIT	P46407 oryctolaqu
20	159.5	61.3	216	1 SOMA_RAT	P01244 rattus norv
21	159.5	61.3	217	1 SOMA_CALSF	O9qka1 galago sene
22	159.5	61.3	217	1 SOMA_NYCPY	O9qmb2 nycticebus
23	156.5	60.2	216	1 SOMA_MUSVI	P19795 mustela vis
24	155.5	59.8	190	1 SOMA_LAMIA	P37885 lama guanac
25	150	57.7	216	1 SOMA_MELGA	P22077 meleagris g
26	148	56.9	191	1 SOMA_CHEMY	P34005 chelonoidis my
27	145	55.8	215	1 SOMA_MONDO	O9q160 monodelphis
28	145	55.8	215	1 SOMA_TRIVU	O62754 trichosurus
29	144	55.4	190	1 SOM1_ACIGU	P26773 acipenser g
30	144	55.4	190	1 SOM2_ACIGU	P26774 acipenser g
31	144	55.4	216	1 SOMA_CHICK	P08998 gallus gall
32	142.5	54.8	217	1 SOMA_BOVIN	P01246 bos taurus
33	142.5	54.8	217	1 SOMA_CEREL	P56437 cervus elap

34 142.5 54.8 217 1 SOMA\_SHEEP  
35 142 54.6 217 1 SOMA\_STRCA  
36 140 53.8 190 1 SOMA\_CROKO  
37 140 53.8 216 1 SOMA\_ANAPL  
38 135.5 52.1 217 1 SOMA\_BUBBU  
39 132 50.8 215 1 SOMA\_RANCA  
40 125 48.1 211 1 SOMA\_LEPOS  
41 122 46.9 214 1 SOMA\_XENLA  
42 116 44.6 183 1 SOMA\_PIGLI  
43 112 43.1 206 1 SOMA\_PROAN  
44 111 42.7 213 1 SOMA\_HUFMA  
45 104 40.0 208 1 SOMB\_XENLA

PC1247 ovis aries  
Q9p9q3 struthio ca  
P55755 crocodylus  
P11228 anas platyr  
O18938 bubalus bub  
P10813 rana catesb  
P79885 lepisosteus  
P12855 xenopus lae  
P34006 prionace gl  
O73848 prototerus  
O73849 bufo marinus  
P12856 xenopus lae

#### ALIGNMENTS

RESULT 1  
SOMA\_HUMAN  
ID SOMA\_HUMAN STANFARD; PRT; 217 AA.  
AC P01241: Q14495, Q16611, Q9H871, Q9H872, Q9H875;  
DI 21-JUL-1986 (Ref. 01, Created)  
DI 01-MAR-1992 (Ref. 21, Last sequence update)  
DI 15-SEP-2003 (Ref. 42, Last annotation update)  
DE Somatotropin precursor (growth hormone) (GH) (GH N) (Pituitary growth hormone) (Growth hormone 1)  
GN GH1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chorda; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=80033447; PubMed=386281;  
RA Roskam W., Rougeon F.  
RT "Molecular cloning and nucleotide sequence of the human growth hormone structural gene."  
RT Nucleic Acids Res. 7:305-320(1979).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=79203293; PubMed=377496;  
RA Martial J.A., Hallewell R.A., Baxter J.D., Goodman H.M.;  
RT "Human growth hormone: complementary DNA cloning and expression in bacteria."  
RT Science 205:602-607(1979).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND POSSIBLE ALTERNATIVE SPLICING.  
RX MEDLINE=82014939; PubMed=6269091;  
RA Denoto F.M., Moore D.D., Goodman H.M.;  
RT "Human growth hormone DNA sequence and mRNA structure: possible alternative splicing."  
RT Nucleic Acids Res. 9:3719-3730(1981).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83182010; PubMed=7169009;  
RA Seeburg P.H.;  
RT "The human growth hormone gene family: nucleotide sequences show recent divergence and predict a new polypeptide hormone."  
RT DNA 1:239-249(1982).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89307277; PubMed=2744760;  
RA Chen E.Y., Liao Y.C., Smith D.H., Herrera-Saldana H.A.,  
RT Gelinis R.E., Seeburg P.H.;  
RT "The human growth hormone locus: nucleotide sequence, biology, and evolution."  
RT Genomics 4:479-497(1989).  
RN [6]  
RP SEQUENCE FROM N.A. (ISOFORM 3).  
RC TISSUE=Pituitary;  
RA Gu J., Huang Q.-H., Li N., Xu S.-H., Han Z.-G., Fu G., Chen Z.;  
RT "A novel gene expressed in human pituitary."  
RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RN RP SEQUENCE FROM N.A. (ISOFORM 4).  
 RC TISSUE-PITUITARY; PubMed:10931946;  
 RX MEDLINE-20402571; PubMed:10931946;  
 RA Hu R.-M., Han Z.-G., Song H.-D., Peng Y. D., Huang Q.-H., Ren S.-X.,  
 RA Gu Y.-J., Huang C.-H., Li Y.-B., Jiang C. L., Fu G., Zhang Q.-H.,  
 RA Gu B.-W., Dai M., Mao Y.-F., Gao G.-F., Xiong R., Ye M., Zhou J.,  
 RA Xu S.-H., Gu J., Shi J.-X., Jin W.-R., Zhang C.-K., Wu T.-M.,  
 RA Huang G.-Y., Chen Z., Chen M.-D., Chen J.-L.,  
 RT "Gene expression profiling in the human hypothalamus-pituitary-adrenal  
 axis and full-length cDNA cloning.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:9543-9548(2000).  
 RN [18]  
 RN RP SEQUENCE OF 1-26 FROM N.A.  
 RX MEDLINE-86137493; PubMed:3912261;  
 RA Gray G.L., Bairdridge J.S., McKewen K.S., Heyneker H.L., Chang C.N.,  
 RT "Periplasmic production of correctly processed human growth hormone in  
 Escherichia coli: natural and bacterial signal sequences are  
 interchangeable.";  
 RL Gene 99:247-264(1985).  
 RN [19]  
 RN RP SEQUENCE OF 27-217  
 RX MEDLINE-69289402; PubMed 5810834;  
 RA Li R.H., Dixon J.S., Liu W.-K.,  
 RT "Human pituitary growth hormone. XIX. The primary structure of the  
 hormone.";  
 RL Arch. Biochem. Biophys. 133:70-91(1965);  
 RN [10]  
 RN RP SEQUENCE OF 27-217, AND REVISIONS.  
 RX MEDLINE-72143935; PubMed:5144027;  
 RA Li C.H., Dixon J.S.,  
 RT "Human pituitary growth hormone. 32. Th. primary structure of the  
 hormone: revision.";  
 RL Arch. Biochem. Biophys. 146:233-236(1971).  
 RN [11]  
 RN RP REVISION  
 RX MEDLINE-73092028; PubMed:4675454;  
 RA Bowley T.A., Dixon J.S., Li C.H.,  
 RT "Sequence comparison of human pituitary growth hormone, human  
 chorionic somatomotropin, and ovine pituitary growth and  
 lactogenic hormones.";  
 RL Int. J. Pept. Protein Res. 4:281-287(1972).  
 RN [12]  
 RN RP SEQUENCE OF 27-61 AND 102-124.  
 RX MEDLINE-71189765; PubMed:5279046;  
 RA Niall H.D.,  
 RT "Revised primary structure for human growth hormone.";  
 RL Nature New Biol. 230:90-91(1971).  
 RN [13]  
 RN RP REVISIONS TO 119-120 AND 157-159.  
 RX MEDLINE 71154968; PubMed:5279528;  
 RA Niall H.D., Hoan M.L., Sauer R., Rosenblum I.Y., Greenwood F.C.,  
 RT "Sequences of pituitary and placental lactogenic and growth hormones:  
 evolution from a primordial peptide by gene reduplication.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 68:866-869(1971).  
 RN [14]  
 RN RP REVISION.  
 RA Niall H.D.,  
 RT "The chemistry of the human lactogenic hormones.";  
 RL (In) Griffiths K. (eds.),  
 RL Prolactin and carcinogenesis. Proc. fourth renovus workshop prolactin,  
 RL pp.13-20. Alpha Omega Alpha Press, Cardiff (1972).  
 RN [15]  
 RN RP SEQUENCE OF 27-79 (ISOFORM 2).  
 RX MEDLINE-81117461; PubMed:7462247;  
 RA Chapman G.E., Rogers K.M., Brittain T., Bradshaw R.A., Bates G.J.,  
 RA Turner C., Cary P.D., Crane-Robinson C.,  
 RT "The 20,000 molecular weight variant of human growth hormone.  
 Preparation and some physical and chemical properties.";  
 RL J. Biol. Chem. 256:2395-2401(1981).  
 RN [16]  
 RN RP SEQUENCE OF 46-80 (ISOFORM 2).  
 RX MEDLINE-80130196; PubMed:7356479;  
 Lewis U.J., Bonewald L.F., Lewis L.J.,  
 RT "The 20,000-dalton variant of human growth hormone: location of the  
 amino acid deletions.";  
 RL Biochem. Biophys. Res. Commun. 92:511-516(1980).  
 RN [17]  
 RN RP DEAMINATION OF GLN-163 AND ASN-178.  
 RX MEDLINE-82052997; PubMed:7028740;  
 RA Lewis U.J., Singh R.N., Bonewald L.F., Seavey B.K.,  
 RT "Altered proteolytic cleavage of human growth hormone as a result of  
 deamidation.";  
 RL J. Biol. Chem. 256:11645-11650(1981).  
 RN [18]  
 RN RP REVIEW.  
 RX MEDLINE-99321812; PubMed:10393484;  
 RA Baumann G.,  
 RT "Growth hormone heterogeneity in human pituitary and plasma.";  
 RL Horm. Res. 51 Suppl. 1:2-6(1999).  
 RN [19]  
 RN RP 3D-STRUCTURE MODELING.  
 RX MEDLINE 88190073; PubMed 447173;  
 RA Cohen P.E., Kuntz L.D.,  
 RT "Prediction of the three dimensional structure of human growth  
 hormone.";  
 RL Proteins 2:162-156(1987).  
 RN [20]  
 RN RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
 RX MEDLINE-92196577; PubMed:1549776;  
 RA de Vos A.M., Uitsch M., Kossiakoff A.A.,  
 RT "Human growth hormone and extracellular domain of its receptor:  
 crystal structure of the complex.";  
 RL Science 255:306-312(1992).  
 RN [21]  
 RN RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).  
 RX MEDLINE-95075462; PubMed:7984244;  
 RA Somers W., Uitsch M., de Vos A.M., Kossiakoff A.A.,  
 RT "The X-ray structure of a growth hormone-prolactin receptor complex.";  
 RL Nature 372:478-481(1994).  
 RN [22]  
 RN RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
 RA Chantalat S., Chirgadze N.Y., Jones N., Korber F., Navaza J.,  
 RA Pavlovsk A.G., Wlodawer A.,  
 RT "The crystal-structure of wild-type growth-hormone at 2.5-A  
 resolution.";  
 RL Protein Pept. Lett. 2:333-340(1995).  
 RN [23]  
 RN RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
 RX MEDLINE-97113023; PubMed:8943276;  
 RA Sundstroem M., Lundqvist T., Rosdin J., Giebel L.B., Millian D.,  
 RA Norstedt G.,  
 RT "Crystal structure of an antagonist mutant of human growth hormone,  
 G120R, in complex with its receptor at 2.9-A resolution.";  
 RL J. Biol. Chem. 271:32197-32203(1996).  
 RN [24]  
 RN RP VARIANT CYS-105.  
 RX MEDLINE-99318093; PubMed:10391209;  
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nimesh J., Ziaugra L.,  
 RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Baley G.O.,  
 RA Lander E.S.,  
 RT "Characterization of single-nucleotide polymorphisms in coding regions  
 of human genes.";  
 RL Nat. Genet. 22:231-238(1999).  
 RN [25]  
 RN RP ERRATUM.  
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nimesh J., Ziaugra L.,  
 RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Baley G.O.,  
 RA Lander E.S.,  
 RL Nat. Genet. 23:373-373(1999).  
 CC -!- FUNCTION: Plays an important role in growth control. Its major  
 CC role in stimulating body growth is to stimulate the liver and  
 CC other tissues to secrete IGF-1. It stimulates both the  
 CC differentiation and proliferation of myoblasts. It also stimulates

Query Match 98.1% Score 255; DB 1; Length 217;  
Best Local Similarity 100.0%; Pred. No. 6.6e-25;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTPLSLRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 49  
DB 27 FPTPLSLRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 74

## RESULT 2

SOMA\_MACMO STANDARD: PRT: 217 AA.  
AC P33093;  
DT 01-OCT-1994 (Rel. 27, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Somatotropin precursor (Growth hormone) (GH) (GH-N) (Pituitary growth hormone) (Growth hormone 1).  
GN GH1.  
OS Macaca mulatta (Rhesus macaque).  
OC Fokaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Macaca.  
OX NCBI\_TaxID:9544;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE:94098724; PubMed:8404617;  
RA Goles T.G., Burning M., Fisher J.M., Fox R.P.D.  
RT "Cloning of four growth hormone/chorion somatomotropin-related complementary deoxyribonucleic acids differentially expressed during pregnancy in the rhesus monkey placenta".  
RL Endocrinology 133:1744-1752(1993).  
RN [2]  
RP SEQUENCE OF 27-217.  
RX MEDLINE:86129460; PubMed:3080959;  
RA Li C.H., Cheng D., Lahm R.W., Stein S.  
RT "The primary structure of monkey pituitary growth hormone".  
RL Arch. Biochem. Biophys. 245:287-291(1986).  
CC -!- FUNCTION: Plays an important role in growth control. Its major role in stimulating body growth is to stimulate the liver and other tissues to secrete IGF-1. It stimulates both the differentiation and proliferation of myoblasts. It also stimulates amino acid uptake and protein synthesis in muscle and other tissues.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: I16556; AAA18842.1;  
PIR: I67410; I67410.  
BSP: p01241; IAXI.  
DR InterPro: IPR001400; Somatotropin.  
DR Pfam: PF00103; hormone; 1.  
DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
KW Hormone; Pituitary; Signal.  
FT SIGNAL: 1 26  
FT CHAIN: 27 217 SOMATOTROPIN.  
FT DISULFID: 79 191 BY SIMILARITY.  
FT DISULFID: 208 215 BY SIMILARITY.  
FT CONFLICT: 100 100 E -> Q (IN REF. 2).  
FT CONFLICT: 179 179 N -> D (IN REF. 2).  
SQ SEQUENCE 217 AA: 24913 MW: 2C5180341EEC46D0 CRC64:

Query Match 98.1% Score 255; DB 1; Length 217;

Best Local Similarity 100.0%; Pred. No. 6.6e-25;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTPLSLRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 49  
DB 27 FPTPLSLRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 74

## RESULT 3

SOMA\_PANTR STANDARD: PRT: 217 AA.  
AC P58756;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Somatotropin precursor (Growth hormone) (GH) (GH-N) (Pituitary growth hormone) (Growth hormone 1).  
GN GH1.  
OS Pan troglodytes (Chimpanzee).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.  
OX NCBI\_TaxID:9598;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Revor A., Esquivel D., Santiago D., Bartera-Saldana H.  
RT "Independent duplication of the growth hormone gene in three Anthropoid lineages".  
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Plays an important role in growth control. Its major role in stimulating body growth is to stimulate the liver and other tissues to secrete IGF-1. It stimulates both the differentiation and proliferation of myoblasts. It also stimulates amino acid uptake and protein synthesis in muscle and other tissues (by similarity).  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AF374232; AA172284.1;  
DR InterPro: IPR001400; Somatotropin.  
DR Pfam: PF00103; hormone; 1.  
DR PRINTS: PR00836; SOMATOTROPIN.  
DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
KW Hormone; Pituitary; Signal.  
FT SIGNAL: 1 26 BY SIMILARITY.  
FT CHAIN: 27 217 SOMATOTROPIN.  
FT DISULFID: 79 191 BY SIMILARITY.  
FT DISULFID: 208 215 BY SIMILARITY.  
SQ SEQUENCE 217 AA: 24843 MW: FEA295E0518674 CRC64:

Query Match 98.1% Score 255; DB 1; Length 217;  
Best Local Similarity 100.0%; Pred. No. 6.6e-25;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 FPTPLSLRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 49  
DB 27 FPTPLSLRFLDNAMLRHRLHQLAFDTYQFEFEAYIPKEQKYSFLONP 74

## RESULT 4

SOMA\_CALJA STANDARD: PRT: 217 AA.  
AC Q9GNB3;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)



DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Somatotropin precursor (Growth hormone).  
 GN GHI.  
 OS *Callithrix jacchus* (Common marmoset).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;  
 OX NCBI\_TaxID=9483;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wallis G.C., Wallis M.;  
 RT "Cloning and characterisation of a putative growth hormone encoding  
 RL gene from the marmoset (*Callithrix jacchus*)."  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Plays an important role in growth control. Its major  
 CC role in stimulating body growth is to stimulate the liver and  
 CC other tissues to secrete IGF-1. It stimulates both the  
 CC differentiation and proliferation of myoblasts. It also stimulates  
 CC amino acid uptake and protein synthesis in muscle and other  
 CC tissues (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: A1295963; CAC03481.1;  
 DR HSSP: P1243; IAZ2.  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone\_1.  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 KW Hormone; Pituitary; Signal.  
 FT SIGNAL 1 26 BY SIMILARITY.  
 FT CHAIN 27 217 SOMATOTROPIN.  
 FT DISULFID 79 191 BY SIMILARITY.  
 FT DISULFID 208 215 BY SIMILARITY.  
 SQ SEQUENCE 217 AA: 24959 MW: 402151A12CE6192 CRC64:  
 -----  
 Query Match 95.8%; Score 249; DB 1; Length 217;  
 Best Local Similarity 97.9%; Pred. No. 3.7e-24;  
 Matches 47; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 2 FTIPLSRFLDNAMLRHLHQLAFDTYOFFEAYIPKEQKYSFLQNP 49  
 DB 27 FPTIPLSRFLDNAMLRHLHQLAFDTYOFFEAYIPKEQKYSFLQNP 74  
 -----  
 RESULT 5  
 SOMA\_SAIBR STANDARD: PRT: 217 AA.  
 ID SOMA\_SAIBR  
 AC P58343;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2004 (Rel. 41, Last annotation update)  
 DE Somatotropin precursor (Growth hormone).  
 GN GHI.  
 OS *Saimiri boliviensis boliviensis* (Bolivian squirrel monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Saimiri.  
 OX NCBI\_TaxID=39432;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=21265430; PubMed=11371582;  
 RA Liu J.C., Makova K.D., Adkins R.M., Gibson J., Li W.H.;  
 RT "Episodic evolution of growth hormone in primates and emergence of the  
 RT species specificity of human growth hormone receptor."  
 RL Mol. Biol. Evol. 18:945-953(2001).

CC -!- FUNCTION: Plays an important role in growth control. Its major  
 CC role in stimulating body growth is to stimulate the liver and  
 CC other tissues to secrete IGF-1. It stimulates both the  
 CC differentiation and proliferation of myoblasts. It also stimulates  
 CC amino acid uptake and protein synthesis in muscle and other  
 CC tissues (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: AF339069; AAK62287.1;  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone\_1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 KW Hormone; Pituitary; Signal.  
 FT SIGNAL 1 26 BY SIMILARITY.  
 FT CHAIN 27 217 SOMATOTROPIN.  
 FT DISULFID 79 191 BY SIMILARITY.  
 FT DISULFID 208 215 BY SIMILARITY.  
 SQ SEQUENCE 217 AA: 24864 MW: 9515289992C529F7 CRC64:  
 -----  
 Query Match 95.8%; Score 249; DB 1; Length 217;  
 Best Local Similarity 97.9%; Pred. No. 3.7e-24;  
 Matches 47; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 2 FTIPLSRFLDNAMLRHLHQLAFDTYOFFEAYIPKEQKYSFLQNP 49  
 DB 27 FPTIPLSRFLDNAMLRHLHQLAFDTYOFFEAYIPKEQKYSFLQNP 74  
 -----  
 RESULT 6  
 SOM2\_PANTR STANDARD: PRT: 217 AA.  
 ID SOM2\_PANTR  
 AC P58757;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Growth hormone variant precursor (GH-V) (Placenta-specific growth  
 DE hormone) (Growth hormone 2).  
 GN GH2.  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.  
 OX NCBI\_TaxID=9598;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Revol A., Esquivel D., Santiago D., Barrera-Saidana H.;  
 RT "Independent duplication of the growth hormone gene in three  
 RT Anthropoidean lineages".  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Plays an important role in growth control. Its major  
 CC role in stimulating body growth is to stimulate the liver and  
 CC other tissues to secrete IGF-1. It stimulates both the  
 CC differentiation and proliferation of myoblasts. It also stimulates  
 CC amino acid uptake and protein synthesis in muscle and other  
 CC tissues.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Expressed in the placenta.  
 CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not recommended. Usage by and for commercial entities requires a license agreement. See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

EMBL: AF374233; AAL72285.1; InterPro: IPR001400; Somatotropin. Pfam: PF00103; hormone; 1. PROSITE: PS00266; SOMATOTROPIN\_1; 1. PROSITE: PS00338; SOMATOTROPIN\_2; 1. Hormone; Placenta; Signal; Glycoprotein. SIGNAL 1 26 BY SIMILARITY. CHAIN 27 217 BY GROWTH HORMONE VARIANT. DISULFID 3 191 BY SIMILARITY. DISULFID 208 215 BY SIMILARITY. SEQUENCE 217 AA; 24990 MW; 15924290756777DE CRC64;

Query Match 60.8%; Score 246; DB 1; Length 217; Best Local Similarity 93.8%; Pred No 160 22; Matches 45; Conserved: 45; Mismatches 1; Gaps 0;

QY 2 PFTITSPSPFNAMIRAHIRHUGAFETVQ FEATPKQKYSFIQNP 49  
II  
DB 27 PFTITSPSPFNAMIRAHIRHUGAFETVQ FEATPKQKYSFIQNP 74

RESULT 7  
SOM2\_HUMAN  
AC P01242; P05987; STANDARD; PRT; 217 AA.  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Growth hormone variant precursor (GH-V) (Placenta-specific growth hormone) (Growth hormone 2).  
GN GH2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RX MEDLINE-8182010; PubMed-7169009;  
RA Seeburg P.H.;  
RT "The human growth hormone gene family: nucleotide sequences show recent divergence and predict a new polypeptide hormone.";  
RL DNA 1:239-249(1982).  
[2]  
RN SEQUENCE FROM N.A. (ISOFORM 1 AND 2).  
RX MEDLINE-88243769; PubMed-3379057;  
RA Cooke N.E., Ray J., Emery J.G., Lieber M.S., and S.A.;  
RT "Two distinct species of human growth hormone variant cRNA in the human placenta predict the expression of a novel growth hormone protein.";  
RL J. Biol. Chem. 263:9001-9006(1988).  
[3]  
RN SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE-89624984; PubMed-2460050;  
RA Ignotz A., Scippo M.L., Frankenreiter F., Hennen G.;  
RT "Cloning and nucleotide sequence of placental hGH-v cDNA.";  
RL Arch. Int. Physiol. Biochim. 96:63-67(1988).  
[4]  
RN SEQUENCE FROM N.A.  
RX MEDLINE-89307277; PubMed-2744760;  
RA Chen E.Y., Liao Y.C., Smith D.H., Baranera-Saldana H.A., Gellinas R.E., Seeburg P.H.;  
RT "The human growth hormone locus: nucleotide sequence, biology, and evolution.";  
RL Genomics 4:479-497(1989).  
[5]  
RN SEQUENCE FROM N.A.  
RC TISSUE: Placenta;  
RX MEDLINE-22388257; PubMed-12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Waqner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Blat N.K., Hopkins R.F., Jordan H., Moore T., Max S.J., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Ronald M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Mazny D.M., Sodergren E., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E., Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16899(2002).  
[5]  
RN REVIEW.  
RX MEDLINE-99321812; PubMed-1049484;  
RA Baumann G.;  
RT "Growth hormone heterogeneity in human pituitary and plasma.";  
RL Horm. Res. 51 Suppl. 1:2-6(1999).  
CC -!- FUNCTION: Plays an important role in growth control. Its major role in stimulating body growth is to stimulate the liver and other tissues to secrete IGF-1. It stimulates both the differentiation and proliferation of myoblasts. It also stimulates amino acid uptake and protein synthesis in muscle and other tissues.  
CC -!- SUBUNIT: Monomer, dimer, trimer, tetramer and pentamer, disulfide-linked or non-covalently associated, in homopolymeric and heteropolymeric combinations. Can also form a complex either with GHBP or with the alpha<sub>2</sub>-macroglobulin complex.  
CC -!- SUBCELLULAR LOCATION: secreted.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1; Synonyms=GH-V1;  
CC IsoId: P01242-1; Sequence=Displayed;  
CC Name=2; Synonyms=GH-V2;  
CC IsoId: P01242-2; Sequence=VSP\_006203;  
CC Note=No experimental confirmation available;  
CC -!- TISSUE SPECIFICITY: Expressed in the placenta.  
CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL: K00470; AAA98619.1;  
DR EMBL: J03756; AAB59547.1;  
DR EMBL: J03756; AAB59549.1;  
DR EMBL: M38451; AAA35891.1;  
DR EMBL: J03671; AAA52552.1;  
DR EMBL: BC020760; AAB20760.1;  
DR PIR: A28072; SIHUV2;  
DR PIR: D32435; SIHUV;  
DR HSSP: P01241; 1A22;  
DR Genew; HGNC:4262; GH2.  
DR MIN: I39240;  
DR GO: GO:0005180; F:peptide hormone; TAS.  
DR InterPro: IPR001400; Somatotropin.  
DR Pfam: PF00103; hormone; 1.  
DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
KW Hormone; Placenta; Signal; Glycoprotein; Alternative splicing;  
KW Polymorphism.  
FT SIGNAL 1 26



RL DNA 1:239-249(1982).  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Placenta, and Uterus;  
 RC MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg K.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schale G.D.,  
 RA Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,  
 RA Shapiro M., Soares M.B., Bonaldo M.P., Casavant T.L., Schenzl P.B.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Kana S.S., Loquellano J., Schumitz J., Myers P.M.,  
 RA Besac S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.D., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gillis R.A.,  
 RA Fahey C., Hefton A., Kettmann M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Yama A., Shengrenko Y., Bouffard G.,  
 RA Blackley K.W., Touchman R.W., Green E.D., Jackson R.,  
 RA Rodriguez A., Grimwood J., Schumitz J., Myers P.M.,  
 RA Butterfield Y.S.N., Krzywicki M.L., Skalsko U., Shattas D.E.,  
 RA Schurch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RA \*Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.\*  
 RT Proc. Natl. Acad. Sci. U.S.A. 93:16899-16904(2002).  
 RL [7].  
 RN SEQUENCE OF 50-217 FROM N.A.  
 RP MEDLINE=7807161; PubMed=593368;  
 RA Shine J., Seeburg P.H., Martial J.A., Baxter J.D., Goodman H.M.,  
 RT \*Construction and analysis of recombinant DNA for human chorionic  
 RT somatomotropin.\*  
 RL Nature 270:494-499(1977).  
 RN [8].  
 RP SEQUENCE OF 27-217.  
 RX MEDLINE=73201971; PubMed=4712450;  
 RA Li C.H., Dixon J.S., Chung D.,  
 RT \*Amino acid sequence of human chorionic somatomotropin.\*  
 RL Arch. Biochem. Biophys. 155:95-110(1973).  
 RN [9].  
 RP SEQUENCE OF 27-117.  
 RX MEDLINE=72016313; PubMed 5280363;  
 RA Sherwood L.M., Handwerger S., McLaurin W.D., Lerner M.,  
 RT \*Amino-acid sequence of human placental lactogen.\*  
 RL Nature New Biol. 233:59-61(1971).  
 RN [10].  
 RP ERRATUM.  
 RA Sherwood L.M., Handwerger S., McLaurin W.D., Lerner M.,  
 RL Nature New Biol. 235:64-64(1972).  
 RN [11].  
 RP INTERCHAIN DISULFIDE BONDS.  
 RX MEDLINE=79473081; PubMed=438159;  
 RA Schneider A.B., Kowalski K., Russell J., Sherwood L.M.,  
 RT \*Identification of the interchain disulfide bonds of dimeric human  
 RL placental lactogen.\*  
 RL J. Biol. Chem. 254:3782-3787(1979).  
 CC -1- FUNCTION: SIMILAR TO THAT OF SOMATOTROPIN.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MISCELLANEOUS: THE SEQUENCE OF CSH1 IS SHOWN.  
 CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: V00573; CAA23836.1; -  
 DR EMBL: J00289; AAA98747.1; -  
 DR EMBL: K02401; AAA52115.1; -  
 DR EMBL: M15894; AAA52116.1; -  
 DR EMBL: J03071; AAA52551.1; -

DR EMBL: J06118; AAA98621.1; -  
 DR EMBL: BC002717; AAH02717.1; -  
 DR EMBL: BC005921; AAH05921.1; -  
 DR EMBL: BC020756; AAH20756.1; -  
 DR PIR: A26449; A26449.  
 DR PIR: C32435; LCHUC.  
 DR HSSP: P01241; 1A22.  
 DR Genes: HGNC:2440; CSH1.  
 DR Genes: HGNC:2441; CSH2.  
 DR MIM: 150200; -  
 DR GO: GO:0007565; P:pregnancy; TAS.  
 DR InterPro: IPR01400; Somatotropin.  
 DR Pfam: PF00103; Hormone; 1.  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 KW Hormone; Placenta; Multidene family; Signal.  
 FT SIGNAL: 1 26  
 FT CHAIN: 27 217  
 FT DISULFID: 75 141  
 FT DISULFID: 208 215  
 FT DISULFID: 208 208  
 FT DISULFID: 215 215  
 FT VARIANT: 3 3  
 FT VARIANT: 104 105  
 FT CONFLICT: 84 84  
 FT CONFLICT: 95 95  
 FT CONFLICT: 116 116  
 FT CONFLICT: 134 136  
 FT CONFLICT: 217 217  
 FT SEQUENCE: 217 AA: 25020 MW: 23580D27A713F431 CRC64;  
 SO  
 Query Match 75.8% Score 197; DB 1; Length 217;  
 Best Local Similarity 80.0%; Pred. No. 130-17;  
 Matches 36; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
 CY 4 TIPLSKLEFDNMLRAHPLHQLAFYQFFEFAYIPKQKYSFLQ 48  
 DB 29 TVPLSRLEFHAMLOAHRAHQLAIDTYQFFETYPKDKYKSLFHQ 73  
 E:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 RESULT 19  
 SOMA\_MOUSE  
 ID SOMA\_MOUSE STANDARD: PRT: 215 AA.  
 AC P06880;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Somatotropin precursor (Growth hormone).  
 GN CSH1 OR GH  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chorda a; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85261358; PubMed=2091252;  
 RA Linzer D.L.H., Talamantes F.;  
 RT \*Nucleotide sequence of mouse prolactin and growth hormone mRNAs and  
 RT expression of these mRNAs during pregnancy.\*  
 RL J. Biol. Chem. 260:9574-9579(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=F27DU; TISSUE: Liver;  
 RX MEDLINE=96194803; PubMed=847448;  
 RA Das P., Meyer L., Seyfert L.-M., Brockmann G., Schwerin M.;  
 RT \*Structure of the growth hormone encoding gene and its promoter in  
 RT mice.\*  
 RL Gene 169:209-213(1996).  
 CC -1- FUNCTION: Plays an important role in growth control. Its major  
 CC role in stimulating body growth is to stimulate the liver and  
 CC other tissues to secrete IGF-1. It stimulates both the  
 CC differentiation and proliferation of myoblasts. It also stimulates





```

CC CC -!- FUNCTION: plays an important role in growth control. Its major
CC CC role in stimulating body growth is to stimulate the liver and
CC CC other tissues to secrete IGF-1. It stimulates both the
CC CC differentiation and proliferation of myoblasts. It also stimulates
CC CC amino acid uptake and protein synthesis in muscle and other
CC CC tissues.
CC CC -!- SUBCELLULAR LOCATION: Secreted.
CC CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC CC at the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/ or
CC CC send an email to license@isb-sib.ch).
CC CC
CC DR EMBL: UZ5974; AAA57294.1;
CC DR EMBL: W43560; AAA56142.1;
CC DR PIR: J04842; p4632.
CC DR HSP: F01246; IRS1.
CC DR InterPro: IPR01490; Somatotropin.
CC DR Prim: PF00103; hormone. 1.
CC DR PROSITE: PS00266; SOMATOTROPIN_1. 1.
CC DR PROSITE: PS00338; SOMATOTROPIN_2. 1.
CC KW Hormone; Pituitary; Signal.
CC FT SIGNAL: 1 26 BY SIMIL Y.
CC FT CHAIN: 27 216 SOMATCTR 1.
CC FT DISULFID: 78 189 BY SIMIL Y.
CC FT DISULFID: 205 214 BY SIMIL Y.
CC FT CONFLICT: 7 7 N -> T (I REF. 2).
CC FT CONFLICT: 26 26 T -> A (I REF. 2).
CC FT CONFLICT: 159 159 G -> A (IN REF. 2).
CC FT CONFLICT: 181 181 L -> P (IN REF. 2).
CC SQ SEQUENCE: 216 AA: 24454 MW: 05820239A7D292C6 CRC64:
CC
CC Query Match: 61.3%; Score 159.5; DB 1; Length 216;
CC Best Local Similarity: 68.1%; Pred.No 6.3e-13;
CC Matches: 32; Conservative: 6; Mismatches: 8; Indels: 1; Gaps: 1;

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:54:30 : Search time 19.1434 Seconds  
(without alignments)  
660,520 Million cell updates/sec

Title: US-09-423-100-1

Perfect score: 260

Sequence: 1 MFPTPLSRFLDNAMLAHR.....QPEEAYIPKESQKSYFLQNP 49

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 3.5

Searched: 830525 seqs, 25805264 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum UB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL-23:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_ordanelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_virus:\*

16: sp\_bacteriophage:\*

17: sp\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result NO.	Score	% Match	Length	DB	ID	Description
1	239	95.8	217	6	Q8WNE0	Q8WNE0 ateles geof
2	228	87.7	245	4	O14644	O14644 homo sapien
3	214	81.9	212	6	Q07368	Q07368 macaca mula
4	213	81.9	217	6	Q07367	Q07367 macaca mula
5	201	77.3	217	6	Q07369	Q07369 macaca mula
6	197	75.8	217	4	O14407	O14407 homo sapien
7	195	75.0	217	6	Q8WNE0	Q8WNE0 ateles geof
8	170	65.4	217	6	O8M174	O8M174 callithrix
9	160.5	61.7	216	11	O70615	O70615 spalax leuc
10	159.5	61.3	52	6	O9TV91	O9TV91 equus cabal
11	159.5	61.3	216	6	O8M173	O8M173 delphinus d
12	159.5	61.3	216	6	O8HYE5	O8HYE5 alluropoda
13	155.5	59.8	216	11	Q9R2C3	Q9R2C3 mus musculus
14	154	59.2	216	11	O9JMK4	O9JMK4 cavia porce
15	152	58.5	178	6	Q95MJ5	Q95MJ5 tarsius ban
16	149.5	57.5	204	6	Q95205	Q95205 oviv aries

17	147	56.5	202	4	O14643	O14643 homo sapien
18	146	56.2	178	6	Q95MJ6	Q95MJ6 tarsius syr
19	144	55.4	190	11	Q9JMK0	Q9JMK0 cavia porce
20	143.5	55.2	192	6	Q9TU21	Q9TU21 capra hircu
21	142.5	54.9	192	6	Q9TOW9	Q9TOW9 bos indicus
22	141	54.2	217	6	O8M175	O8M175 callithrix
23	138.5	53.3	217	6	Q28957	Q28957 sus scrofa
24	138.5	53.3	217	6	Q9BEC0	Q9BEC0 traquilus ja
25	138.5	53.3	217	6	Q9BEB9	Q9BEB9 traquilus ja
26	137	52.7	40	6	Q9TR19	Q9TR19 macropus ru
27	132	50.8	218	13	Q9PU72	Q9PU72 cynops pyrri
28	128	49.2	195	13	Q91386	Q91386 amia calva
29	123.5	47.5	143	6	Q95240	Q95240 canis fami
30	122	46.9	199	4	O14406	O14406 homo sapien
31	119	45.8	63	13	Q8QC85	Q8QC85 anser anser
32	107	41.2	53	6	O19034	O19034 oviv aries
33	104	40.0	55	6	Q46474	Q46474 telis silve
34	101	38.8	167	4	P78451	P78451 homo sapien
35	101	38.8	200	13	Q8QFM8	Q8QFM8 ciarias bat
36	99.5	38.3	145	6	Q9BDR4	Q9BDR4 galago cras
37	99	38.3	200	13	Q8QFN2	Q8QFN2 heteropneus
38	99	38.3	200	13	Q8QFN2	Q8QFN2 ciarias gar
39	96	36.9	50	13	Q9PR85	Q9PR85 ciarias gar
40	95	36.5	200	13	O8AV82	O8AV82 silurus aso
41	90	34.6	188	13	Q90283	Q90283 carassius a
42	87	33.5	188	13	O8UVF2	O8UVF2 cirrhinus m
43	87	33.5	188	13	Q90W26	Q90W26 carassius a
44	87	33.5	188	13	Q90W27	Q90W27 carassius a
45	87	33.5	188	13	Q98SR7	Q98SR7 cyprinus ca

## ALIGNMENTS

RESULT 1

Q8WNE0 PRELIMINARY: PRT: 217 AA.

AC Q8WNE0: 01-MAR-2002 (TRENBLrel. 20, Created)

DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)

DI 01-MAR-2003 (TRENBLrel. 23, Last annotation update)

DE Growth hormone.

GN GH-N.

OS Ateles geoffroyi (Black-handed spider monkey).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catartida; Cebidae; Ateles.

ON NCBI\_TaxID=9509;

RN [1]

RP SEQUENCE FROM N.A.

RA Revol. A., Esquivel D., Santiago D., Barrera-Saldana H.,

RT "Independent duplication of the growth hormone gene in three

RT Anthropoledean lineages."

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF374234; AAL72286.1; ...

DR InterPro: IPR001400; Somatotropin.

DR Pfam: PF00103; hormone; 1

DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.

DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.

SQ SEQUENCE 217 AA: 24894 MW: 425829FF41EEAAE6 CRC64:

Query Match: 95.8%; Score 249; ID 6; Length 217;

Best Local Similarity: 97.9%; Pred. No. 1.2e-25;

Matches: 47; Conservative: 0; Mismatches: 1; Indels: 0; Gaps: 0;

QY 2 FPTPLSRFLDNAMLAHR.HOLAHDYQEFEEAYIPKESQKSYFLQNP 49

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 27 FPTPLSRLLDNAMLAHR.HOLAHDYQEFEEAYIPKESQKSYFLQNP 74

RESULT 2

O14644 PRELIMINARY: PRT: 245 AA.

ID O14644

AC O14644;



**Qy**            3 PTPLSRFLFONAMLRRAHRLHQLAFTDYQEFEAYIPKEQKYSFLQNP 49  
               | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
**Dd**            23 PSVPLSRFLFDHAMIQARHLHQLAFTDYQEFEAYIPKEKKHSLMENP 69

GH-V.  
GN Ateles Geoffroyi (Black-handed spider monkey).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Atelinae; Ateles.  
OX NCBI\_TaxID=9509;  
KN [1]  
RP SEQUENCE FROM N.A.  
RA Revol A., Esquivel D., Santiago B., Barrera-Saldana H.;  
RT \*Independent duplication of the growth hormone gene in three  
RL Anthropoidcan lineages.\*;  
DR Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF374235; AAL72287.1;  
DR InterPro: IPR001400; Somatotropin.  
DR Pfam: PF00103; hormone; 1.  
DR PRINTS: PR00836; SOMATOTROPIN.  
DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
LR PROSITE: PS00348; SOMATOTROPIN\_2; 1.  
SQ SEQUENCE 217 AA: 45293 MW: 74175 ALKASGSGRECRCKE4;

Query Match 75.0%; Score 196; ID # Length 217;  
Best Local Similarity 77.1%; Pred. No. 2 Sc-Id;  
Matches 37; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OY 2 FPIPLPSRI DNAMLKARHLHGLAFDTYORFEAYIPKEUKYSFLQNP 49  
II IIIIII : IIII:::IIIIII I: IIIII II:II  
Db 27 FPIPLPSRLGDAMLKAHQHVAFDYUHLSENCIPKKKYFFELNPP 74

RESULT 8  
OBM174 PRELIMINARY: PRT: 217 AA.

ID OBM174 AC  
AC OBM174:  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Growth hormone-like protein 6 precursor.  
GN GHLP6.  
OS Callithrix jacchus (Common marmoset).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callithrix.  
OX NCBI\_TaxID=9483;  
KN [1]  
RP SEQUENCE FROM N.A.  
RA Wallis G.C., Wallis M.;  
RT \*Characterisation of the GH gene cluster in a New-world monkey, the  
RT marmoset (Callithrix jacchus)\*;  
RL J. Mol. Endocrinol. 0:0-(2002).  
DR EMBL; AJ489811; CAD34012.1;  
DR InterPro: IPR001400; Somatotropin.  
DR Pfam: PF00103; hormone; 1.  
DR PRINTS: PR00836; SOMATOTROPIN.  
DR PROSITE: PS00348; SOMATOTROPIN\_2; 1.  
KW Signal.  
FT SIGNAL: 1 26 POTENTIAL.  
ET CHAIN: 27 217 GROWTH HORMONE LIKE PROTEIN 6  
SQ SEQUENCE 217 AA: 25177 MW: 5ECF14B79B27BF1A CRCK64;

Query Match 65.4%; Score 170; ID # Length 217;  
Best Local Similarity 68.1%; Pred. No. 6 Sc-Id;  
Matches 32; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

OY 3 PTIPLSKFDNAMLKARHLHGLAFDTYORFEAYIPKEUKYSFLQNP 49  
I IIII::: IIII:::IIIIII I: IIIII II:II  
Db 28 PRIPLSKFGDMLRKARQLHILALETYYREEFKNCVPEUKYFFELNPP 74

RESULT 9  
O70615 PRELIMINARY: PRT: 216 AA.

ID O70615 AC  
AC O70615:  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)



```

Q9R2C3
ID Q9R2C3 PRELIMINARY: PRT: 216 AA.
AC Q9R2C3
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Growth hormone.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Nguyen T.N.K., Lieberhaber S.A.
RT "Mouse Growth Hormone Locus: Nucleotide Sequence and Phylogenetic
RT Analyses."
RL EMBL: U34462; AAC99968.1;
DR HSSP: P01241; IAXI.
DR InterPro: IPR001400; Somatotropin
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
SQ SEQUENCE 216 AA: 24682 MW: 455618 CRC64:

Query Match 59.8%; Score 15; DB 11; Length 216;
Best Local Similarity 66.0%; Pred. No. 6e-13;
Matches 31; Conservative 6; Mismatches 9; Indels 1; Gaps 1;

QY 2 FPTPLSLRFDNAMLRAHRLHQLAFDTYQEEAYIPKEQKYSFLON 48
Db 27 FPMPLSLRFDNAMLRAHRLHQLAADIYKLEAYIPKEQKYSFLON 72

RESULT 14
Q9JRM4
ID Q9JRM4 PRELIMINARY: PRT: 216 AA.
AC Q9JRM4
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Growth hormone precursor.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricomorpha; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RA Odorico D.M., Fuller P.J., Herington A.C.
RT "Cloning and sequence of guinea pig growth hormone (GH).";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBCLS databases.
DR EMBL: AF234853; AAF36409.1;
DR HSSP: P01241; IAXI.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
KW Signal.
FT SIGNAL: 1 26 POTENTIAL
FT CHAIN: 27 216 GROWTH HORMONE.
SQ SEQUENCE 216 AA: 24822 MW: 459968 CRC64:

Query Match 59.2%; Score 154; DB 11; Length 216;
Best Local Similarity 67.4%; Pred. No. 8.9e-13;
Matches 29; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 2 FPTPLSLRFDNAMLRAHRLHQLAFDTYQEEAYIPKEQKYS 44
Db 27 FPMPLSLRFDNAMLRAHRLHQLAADIYKLEAYIPKEQKYS 69

```

```

RESULT 15
Q95MJ5
ID Q95MJ5 PRELIMINARY: PRT: 178 AA.
AC Q95MJ5
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Growth hormone (Fragment).
OS Tarsius bancanus (Western tarsier) (Horsfield's tarsier).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Tarsiidae; Tarsiidae; Tarsiidae.
OX NCBI_TaxID=9477;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-21265430; PubMed-11371562;
RA Liu J.C., Makova K.D., Adkins R.M., Gibson S., Li W.H.
RT "Episodic Evolution of Growth Hormone in Primates and Emergence of the
RT Species Specificity of Human Growth Hormone Receptor."
RL EMBL: AF390611; AB945953(2001);
DR HSSP: P01241; IAXI.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
FT NON_TER 1 178
FT NON_TER 178 178
SQ SEQUENCE 178 AA: 20038 MW: 4678087512F9H7FC CRC64:

Query Match 58.5%; Score 152; DB 6; Length 178;
Best Local Similarity 65.1%; Pred. No. 1.4e-12;
Matches 28; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 FPTPLSLRFDNAMLRAHRLHQLAFDTYQEEAYIPKEQKYS 44
Db 6 FPSVPLSKLFANAVLRAHRLHQLAADIYKLEAYIPKEQKYS 50

Search completed: September 15, 2003, 12:03:27
Job time : 22.1434 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:44:15 : Search time 64.6308 Seconds  
(without alignments)  
225,942 Million cell updates/sec

Title: US-09-423-100-2

Perfect score: 470

Sequence: 1 MFPTIPLSRIFDNAMLAHRR.....NLELRISLLILSLWLEPVQ 52

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1167863 seqs, 154274573 residues

Total number of hits satisfying chosen parameters: 1000000

Minimum hit seq length: 5

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

1:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA2002.DAT.*
24:	/SIDS1/qcadata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Query Match	Length DB	ID	Description
1	470	100.0	92	20	AA19856 Human growth hormo
2	470	100.0	134	20	AA192265 Human anti-angioten
3	470	100.0	150	20	AA192861 Chimeric protein,
4	465	98.9	140	10	AA191041 Human growth hormo
5	465	98.9	192	10	AA190129 Human growth hormo
6	465	98.9	192	20	AA192264 Human anti-angioten
7	465	98.9	261	10	AA191299 Human nerve growth
8	465	98.9	262	12	AA191740 Human growth hormo
9	465	98.9	310	11	AA193255 Fusion protein of

10	462	98.3	144	11	AA195313 Segment of B-cell
11	462	98.3	262	7	AA191033 Human beta-nerve g
12	460	97.9	138	9	AA191226 Sequence of protei
13	460	97.9	191	18	AA192010 Primary amino acid
14	460	97.9	191	20	AA195809 Natural human 22kD
15	460	97.9	191	20	AA194396 Mutant human 22kDa
16	460	97.9	191	20	AA194397 Human growth hormo
17	460	97.9	191	21	AA1978425 Human growth hormo
18	460	97.9	191	22	AA197485 Human growth hormo
19	460	97.9	191	22	AA197486 Human growth hormo
20	460	97.9	191	23	AB194859 Human growth hormo
21	460	97.9	191	23	AB194860 Human growth hormo
22	460	97.9	191	23	AB194861 Human growth hormo
23	460	97.9	191	23	AB194862 Human growth hormo
24	460	97.9	191	23	AB194863 Human growth hormo
25	460	97.9	191	23	AB194864 Human growth hormo
26	460	97.9	191	23	AB194907 Human growth hormo
27	460	97.9	191	23	AB194908 Human growth hormo
28	460	97.9	191	23	AB194909 Human growth hormo
29	460	97.9	191	23	AB194910 Human growth hormo
30	460	97.9	191	23	AB194911 Human growth hormo
31	460	97.9	191	23	AB194912 Human growth hormo
32	460	97.9	191	23	AB194913 Human growth hormo
33	460	97.9	191	23	AB194914 Human growth hormo
34	460	97.9	191	23	AB194915 Human growth hormo
35	460	97.9	191	23	AB194916 Human growth hormo
36	460	97.9	191	2	AB194917 Human growth hormo
37	460	97.9	191	2	AB194918 Human growth hormo
38	460	97.9	191	23	AB194919 Human growth hormo
39	460	97.9	191	23	AB194920 Human growth hormo
40	460	97.9	191	23	AB194921 Human growth hormo
41	460	97.9	191	23	AB194922 Human growth hormo
42	460	97.9	191	23	AB194923 Human growth hormo
43	460	97.9	191	23	AB194924 Human growth hormo
44	460	97.9	191	23	AB194925 Human growth hormo
45	460	97.9	191	23	AB194926 Human growth hormo

ALIGNMENTS

RESULT 1  
AA192856  
ID AA192856 standard; protein: 92 AA.  
XX  
AC AA192856:  
XX  
DT 19-JAN-2000 (first entry)  
XX  
DE Human growth hormone (HGH) N terminal fragment #2.  
XX  
KW Growth hormone; chaperone; intramolecular; insulin; precursor;  
KW folding; conformation; chimeric protein; cleavable; recombinant;  
KW production; yield.  
XX  
OS Homo sapiens.  
XX  
PN WO/95/0102 A1.  
XX  
PD 07-OCT-1999.  
XX  
PF 31-MAR 1998; 98WO/000052.  
XX  
PR 31-MAR-1998; 98WO/000052.  
XX  
PA (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.  
XX  
PI Gan 2;  
XX  
DR WPI: 1999-610839/52.  
XX  
PT New chimeric proteins containing human growth hormone fragment, used particularly for the production of human insulin.



PN W09950302-A1.  
 XX 07-OCT-1999.  
 XX 31 MAR-1998: 98WO-CN000052.  
 XX 31 MAR-1998: 98WO-CN000052.  
 XX (TUNG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.  
 XX Gun Z.  
 XX WPI: 1999-610839/52.  
 XX New chimeric proteins containing human growth hormone fragment, used  
 PT particularly for the production of human insulin.  
 XX Claim 14: Line 40-51: 46pp: English.  
 XX This sequence represents a chimeric protein, which contains an  
 CC N-terminal fragment of human growth hormone (hGH) of the sequence given  
 CC in AA42856, a cleavable peptide linker (AA42857), and a human insulin  
 CC precursor comprising insulin A and B chains (AA42859). The hGH portion  
 CC of the chimeric protein acts as an intramolecular chaperone (IMC) for  
 CC the insulin precursor, enabling it to fold correctly. The cleavable  
 CC peptide linker has a C-terminal Arg residue which enables the hGH portion  
 CC of the chimeric protein to be removed after folding has taken place.  
 CC Production of recombinant human insulin: an hGH-proinsulin chimERIC  
 CC protein can provide human insulin with correctly linked cysteine bridges  
 CC with fewer necessary procedural steps, and hence resulting in a higher  
 CC yield of human insulin. The IMC sequences not only protect insulin  
 CC sequences from intracellular degradation by a microorganism host, but  
 CC also promote the folding of the fused insulin precursor, facilitate the  
 CC solubility of the fusion protein and decrease the intermolecular  
 CC interactions among the fusion proteins, thus allowing folding of the  
 CC fused insulin precursor at commercially useful high concentrations. The  
 CC procedural steps of cyanogen bromide cleavage, oxidative sulphitolysis  
 CC and related purification steps can thus be eliminated, along with the use  
 CC of high concentrations of mercaptan or the use of hydrophobic absorbent  
 CC resins.  
 XX  
 XX Sequence 150 AA:  
 SQ Query Match 100.0% Score 476; DB 20; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 36-39;  
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MEFTTPLSRLEFONAMLRHRLHQLAFDTYQFEFAYIPKEOKYSLQNPTSLSPSESIP 60  
 DB 1 MEFTTPLSRLEFONAMLRHRLHQLAFDTYQFEFAYIPKEOKYSLQNPTSLSPSESIP 60  
 QY 61 TPSNREETOQKSNLELLRLISLLLIQSWLEPVQ 92  
 DB 61 TPSNREETOQKSNLELLRLISLLLIQSWLEPVQ 92  
 RESULT 4  
 AAF91041  
 ID AAF91041 standard; protein: 140 AA.  
 XX  
 XX AAF91041:  
 XX 14-OCT-1989 (first entry).  
 XX Human growth hormone segment.  
 XX Human growth hormone; fusion protein; thrombin;  
 KW geriatric dementia; nervous disorders; human nerve factor.  
 XX Homo sapiens (human).  
 OS EP329175-A.  
 PN

PD 23-AUG-1989.  
 XX 17-FEB-1989: 89EP-0102795.  
 XX 19-FEB-1988: 88JP-0035042.  
 XX (TOXI) TOSOH CORP.  
 XX Ohtesuka F.  
 XX WPI: 1989-243092/34.  
 XX New human nerve growth factor gene encoding fusion protein  
 PT - having cleavage site for thrombin, useful for treating geriatric  
 PT dementia, etc.  
 XX Disclosure: page 21: 49pp: English.  
 XX Human growth hormone segment, used at the N-terminal of a fusion  
 CC protein, which contains a thrombin recognition site, and human beta nerve  
 CC growth factor (beta-NGF) at the C-terminal. Beta-NGF can be used to  
 CC control geriatric dementia and other nervous disorders, and can be  
 CC released from the fusion protein by incubation with thrombin (see  
 CC AAN90577-8, AAP91034, AAP91299).  
 XX Sequence 140 AA:  
 SQ Query Match 98.9% Score 465; DB 10; Length 140;  
 Best Local Similarity 98.9%; Pred. No. 36-39;  
 Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MEFTTPLSRLEFONAMLRHRLHQLAFDTYQFEFAYIPKEOKYSLQNPTSLSPSESIP 60  
 DB 1 MEFTTPLSRLEFONAMLRHRLHQLAFDTYQFEFAYIPKEOKYSLQNPTSLSPSESIP 60  
 QY 61 TPSNREETOQKSNLELLRLISLLLIQSWLEPVQ 92  
 DB 61 TPSNREETOQKSNLELLRLISLLLIQSWLEPVQ 92  
 RESULT 5  
 AAF90129  
 ID AAF90129 standard; protein: 192 AA.  
 XX  
 XX AAF90129:  
 XX 25-MAR-2003 (updated)  
 DT 06-FEB-1996 (revised)  
 DT 01-NOV-1989 (first entry)  
 XX Human growth hormone.  
 XX Human growth hormone; fusion protein; recombinant  
 KW vector.  
 XX Homo sapiens (Human).  
 OS JP01144981-A.  
 PN 07-JUN-1989.  
 XX 02-DEC-1987: 87JP-0304937.  
 PF 02-DEC-1987: 87JP-0304937.  
 XX (WAKT) WAKUNAGA SEIYAKU KK.  
 PA WPI: 1989-209284/29.  
 DR N-PSDB: AAN90269.  
 XX Recombinant vector contg. fusion protein - consisting of human  
 PT growth hormone or deriv. ligated to foreign protein, for stability  
 PT and high yield.





```

XX DR WPI: 1989-0243092/34.
XX PT New human nerve growth factor gene encoding fusion protein
XX PT - having cleavage site for thrombin, useful for treating geriatric
XX PT dementia, etc.
XX PS
XX PS Claim 36; page 31-32; 38pp; English.
XX CC Fusion protein consisting of human growth hormone at the
XX CC N-terminal end (1st region), a 3 amino acid sequence representing
XX CC thrombin recognition site, and human beta nerve growth factor (beta-NGF)
XX CC at the C-terminal. Beta-NGF can be used to control geriatric dementia
XX CC and other nervous disorders, and can be released from the fusion
XX CC protein by incubation with thrombin (see AAN95577-8, AAP91034,
XX CC AAP91041).
XX SQ Sequence 262 AA.
XX Query Match 98.9%; Score 465; DB 10; Length 262;
XX Best Local Similarity 98.9%; Pred. NO. 49;
XX Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 MFPTIPLSKLFDNAMLRHRLHQLAFDTYQFEEAYIPKEQKYSFLONPOTSLSEFSESIP 60
XX Db 1 MFPTIPLSKLFDNAMLRHRLHQLAFDTYQFEEAYIPKEQKYSFLONPOTSLSEFSESIP 60
XX QY 61 TPSNREETQOKSNLELLRISILLIQTOSWLEPVQ 92
XX Db 61 TPSNREETQOKSNLELLRISILLIQTOSWLEPVQ 92
XX
XX RESULT 8
XX AAR11740
XX ID AAR11740 standard; Protein: 262 AA.
XX AC AAR11740;
XX DT 25-MAR-2003 (updated)
XX DT 25-JUN-1991 (first entry)
XX DE Human growth hormone/human nerve growth factor beta fusion protein.
XX KW hGH; hNGF; nervous system diseases; dementia.
XX OS Homo sapiens.
XX XX JP03067548 A.
XX PD 22-MAR-1991.
XX PF 07-AUG-1989; 88JP-0202835.
XX PR 07-AUG-1989; 88JP-0202835.
XX PA (TOYJ) TOSOH CORP.
XX The protein is produced by fusing DNA encoding BDNF (1L-) with DNA
XX encoding hGF-2 (1L-5) and ligating the product into an expression vector
XX See also AAR05311 and AAR05314.
XX SQ Sequence 310 AA;
XX Query Match 98.9%; Score 465; DB 11; Length 310;
XX Best Local Similarity 98.9%; Pred. NO. 73e-39;
XX Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 MFPTIPLSKLFDNAMLRHRLHQLAFDTYQFEEAYIPKEQKYSFLONPOTSLSEFSESIP 60
XX Db 1 MFPTIPLSKLFDNAMLRHRLHQLAFDTYQFEEAYIPKEQKYSFLONPOTSLSEFSESIP 60
XX QY 61 TPSNREETQOKSNLELLRISILLIQTOSWLEPVQ 92
XX Db 61 TPSNREETQOKSNLELLRISILLIQTOSWLEPVQ 92
XX
XX RESULT 10
XX AAR05313
XX ID AAR05313 standard; protein: 144 AA.
XX

```

```

XX SQ Sequence 262 AA.
XX Query Match 98.9%; Score 465; DB 12; Length 262;
XX Best Local Similarity 98.9%; Pred. NO. 4e-39;
XX Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 MFPTIPLSKLFDNAMLRHRLHQLAFDTYQFEEAYIPKEQKYSFLONPOTSLSEFSESIP 60
XX Db 1 MFPTIPLSKLFDNAMLRHRLHQLAFDTYQFEEAYIPKEQKYSFLONPOTSLSEFSESIP 60
XX QY 61 TPSNREETQOKSNLELLRISILLIQTOSWLEPVQ 92
XX Db 61 TPSNREETQOKSNLELLRISILLIQTOSWLEPVQ 92
XX
XX RESULT 9
XX AAR05255
XX ID AAR05255 standard; protein: 419 AA.
XX AC AAR05255;
XX DT 19-JUL-1990 (first entry)
XX DE Fusion protein of B-cell stimulatory factor-2 and B-cell
XX DE differentiation factor.
XX KW B-cell stimulatory factor-2 interleukin-6; B cell differentiation;
XX KW interleukin-6; fusion prote ...
XX OS Homo sapiens.
XX XX JP02013375-A.
XX PD 17-JAN-1990.
XX PF 01-JUL-1988; 88JP-0162556.
XX PR 01-JUL-1988; 88JP-0162556.
XX PA (TOYJ) TOSOH CORP.
XX WPI: 1990-062207/05.
XX DR N-PSDB; AAQ02028.
XX The protein is produced by fusing DNA encoding BDNF (1L-) with DNA
XX encoding hGF-2 (1L-5) and ligating the product into an expression vector
XX See also AAR05311 and AAR05314.
XX SQ Sequence 310 AA;
XX Query Match 98.9%; Score 465; DB 11; Length 310;
XX Best Local Similarity 98.9%; Pred. NO. 7.3e-39;
XX Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 MFPTIPLSKLFDNAMLRHRLHQLAFDTYQFEEAYIPKEQKYSFLONPOTSLSEFSESIP 60
XX Db 1 MFPTIPLSKLFDNAMLRHRLHQLAFDTYQFEEAYIPKEQKYSFLONPOTSLSEFSESIP 60
XX QY 61 TPSNREETQOKSNLELLRISILLIQTOSWLEPVQ 92
XX Db 61 TPSNREETQOKSNLELLRISILLIQTOSWLEPVQ 92
XX
XX RESULT 10
XX AAR05313
XX ID AAR05313 standard; protein: 144 AA.
XX

```

```

AC AAR05313;
XX
XX DT 19-JUL-1990 (first entry)
XX
XX DB Segment of B-cell stimulatory factor-2 (IL-5).
XX
XX KW B-cell stimulatory factor-2; interleukin-5.
XX
XX OS Homo sapiens.
XX
XX PN JP02013375-A.
XX
XX PD 17-JAN-1990.
XX
XX PF 01 JUL-1988; 88JP-0162556.
XX
XX PR 01-JUL-1988; 88JP-0162556.
XX
XX PA (TOYU ) TOYO SODA MFG CO LTD.
XX
XX KW WPI: 1990 062207/09.
XX
XX DB N-PSDB: AAQ02028.
XX
XX PT Prepn. of human B-cell differentiation factor - from specified DNA
XX sequence segment, by recombinant DNA technique, gives protein of
XX specified amino acid sequence.
XX
XX PS Disclosure: Page 9; 17pp; Japanese.
XX
XX CC The sequence encoding this protein can be fused with DNA encoding B-cell
XX differentiation factor (IL-6) and ligated into an expression vector for
XX prodn. of a fusion protein.
XX
XX CC See also AAR05311.
XX
XX SQ Sequence 144 AA;

Query Match 98.38; Score 462; DB 11; Length 144;
Best local similarity 97.88; Prod. No. 1 to 39;
Matches 90; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MEPTIPLSRLFDNAMLRAHRLHOLAFDTYQEFEEAYIPKEQKYSFLONPQTSLSSESIP 60
DB 1 MEPTIPLSRLFDNAMLRAHRLHOLAFDTYQEFEEAYIPKEQKYSFLONPQTSLSSESIP 60
QY 61 TFSNREETOQKSNLELLRISILLIQSWLEPVQ 92
DB 61 TFSNREETOQKSNLELLRISILLIQSWLEPVQ 92

RESULT 12
AAP61033
ID AAP61033 standard; Protein: 262 AA
XX
XX AC AAP61033;
XX
XX DT 25-MAR-2003 (updated)
XX
XX DT 20-NOV-1990 (first entry)
XX
XX DE Sequence of protein with somatomedin-like activity.
XX
XX KW Growth hormone.
XX
XX OS Synthetic.
XX
XX PN JP63167798-A.
XX
XX PD 11-JUL-1988.
XX
XX PF 29-DEC-1986; 86JP-0310177.
XX
XX PR 29-DEC 1986; 86JP-0310177.
XX
XX PA (TOYU ) TOYO SODA MFG CO LTD.
XX
XX DB WPI: 1988-232632/33.
XX
XX DB N-PSDB: AAN81605.
XX
XX PT Polypeptide with somatomedin like activity .
XX by culturing bacterium transformed by plasmid conta. gene
XX segment with specified DNA sequence.
XX
XX PS Claim 2(1); Page 609; 9pp; Japanese.
XX
XX CC The polypeptide (AAP81226) with somatomedin-like activity and the DNA
XX (AAN81605) encoding it are claimed. A Met residual gp. may be added to
XX the N-terminal. The polypeptide acts on the bone structure of mammals,
XX including humans, to promote bone growth. The polypeptide has high
XX production rate and is easily extracted from bacterial culture medium
XX and refined for use as a bone growth accelerator.
XX (updated on 25-MAR-2003 to correct PA field.)
XX

```



```
XX 06-APR-1999.
XX
XX 22-SEP-1997; 97JP-0275277.
XX
XX 22-SEP-1997; 97JP-0275277.
XX
XX (SUMU ) SUMITOMO SEIYAKU KK.
XX
XX WP1; 1999-283567/24.
XX
XX A human growth hormone mutant - with equivalent activity to natural
XX human growth hormone
XX
XX Example 1; Page 5-6; 10pp; Japanese.
XX
XX The present invention describes a human growth hormone mutant in which
XX the 134th, Ala and the 135th Thr are replaced respectively by Asp and Pro
XX in the 1st to the 191st amino acid sequence of natural type human 22 kDa
XX growth hormone (hGH) and which has a resistance against decomposition by
XX thrombin. The present sequence represents the natural hGH. Also
XX described are: (1) a hGH mutant in which the 134th Arg, the 135th Thr
XX and the 140th Lys are replaced respectively by Asp, Pro and Ala in the
XX amino acid sequence of natural type hGH and which has a resistance
XX against decomposition by thrombin and plasmin; and (2) a drug
XX preparation containing the above hGH mutant as the active component.
XX The mutant hGH shows an activity approximately equivalent to that of
XX natural type hGH and shows a high stability in blood and body fluid.
XX
XX Sequence 191 AA:
SQ
Query Match 97.9%; Score 460; DB 20; Length 191;
Best Local Similarity 98.9%; Pred. No. 1 3e-38;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 FPTPLSRFLDNAMLRHRLHQIAFDIYQFFEEAYIPKEQKYSFLQNPOTSLSFSISPT 61
DB 1 FPTPLSRFLDNAMLRHRLHQIAFDIYQFFEEAYIPKEQKYSFLQNPOTSLSFSISPT 60
OY 62 PSNREFTQOKSNLELLKISLLILQSWLEPVQ 92
DB 61 PSNREFTQOKSNLELLKISLLILQSWLEPVQ 91
Search completed: September 15, 2003, 12:00:57
Job time : 65.6309 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:57:40 : Search time 21.1039 Seconds  
(without alignments)  
184.449 Million cell updates/sec

Title: US-09-423-100-2

Perfect score: 470

Sequence: 1 MFPTPLSRLEFDNAMLRAHR.....NEELLRISILLIQSWLEPVQ 92

Scoring table: BLASTSUM62

Gapop 10.0 : Gapext 0.5

Searched: (28717 seqs, 421,458 residues)

Total number of hits satisfying chosen parameters: 32717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, AA:\*

1: /cqn2\_6/ptodata/1/iaa/5A.COMB.pep:\*

2: /cqn2\_6/ptodata/1/iaa/5B.COMB.pep:\*

3: /cqn2\_6/ptodata/1/iaa/6A.COMB.pep:\*

4: /cqn2\_6/ptodata/1/iaa/6B.COMB.pep:\*

5: /cqn2\_6/ptodata/1/iaa/PTUS.COMB.pep:\*

6: /cqn2\_6/ptodata/1/iaa/ba-xfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	465	98.9	192	1	Sequence 1, Appli
2	460	97.9	191	4	Sequence 5, Appli
3	460	97.9	194	2	Sequence 4, Appli
4	460	97.9	194	3	Sequence 4, Appli
5	460	97.9	217	3	Sequence 10, Appli
6	460	97.9	217	3	Sequence 10, Appli
7	460	97.9	217	3	Sequence 10, Appli
8	460	97.9	217	3	Sequence 11, Appli
9	460	97.9	217	3	Sequence 11, Appli
10	460	97.9	241	4	Sequence 25, Appli
11	460	97.9	245	4	Sequence 25, Appli
12	460	97.9	274	3	Sequence 66, Appli
13	460	97.9	360	3	Sequence 73, Appli
14	454	96.6	191	4	Sequence 1, Appli
15	454	96.6	217	1	Sequence 4, Appli
16	454	96.6	217	1	Sequence 4, Appli
17	454	96.6	217	2	Sequence 51, Appli
18	454	96.6	217	2	Sequence 51, Appli
19	454	96.6	217	4	Sequence 4, Appli
20	453	96.4	400	4	Sequence 36, Appli
21	453	96.4	401	4	Sequence 37, Appli
22	447	95.1	191	3	Sequence 18, Appli
23	445	94.7	191	3	Sequence 16, Appli
24	445	94.7	191	3	Sequence 20, Appli
25	364.5	77.6	176	3	Sequence 1, Appli
26	364.5	77.6	176	4	Sequence 1, Appli
27	358.5	76.3	176	3	Sequence 2, Appli

28 358.5 76.3 176 4 US-08-990-774-2 Sequence 2, Appli

29 340 72.3 168 6 5424199-3 Patent No. 5424199

30 333.5 71.0 198 1 US-08-187-756C-5 Sequence 5, Appli

31 333.5 71.0 198 2 US-08-710-324A-5 Sequence 5, Appli

32 333.5 71.0 198 4 US-09-411-657-5 Sequence 8, Appli

33 306.5 65.2 191 1 US-08-468-824-8 Sequence 4, Appli

34 304.5 64.8 191 1 US-07-963-331D-4 Sequence 1, Appli

35 302.5 64.4 216 2 US-09-105-651-3 Sequence 2, Appli

36 301.5 64.1 190 1 US-08-388-267C-2 Sequence 2, Appli

37 301.5 64.1 190 4 US-09-277-720-2 Sequence 2, Appli

38 301.5 64.1 191 6 521018C-1 Patent No. 5210180

39 301.5 64.1 193 1 US-07-621-187C-2 Sequence 2, Appli

40 301.5 64.1 193 1 US-08-363-982-2 Sequence 2, Appli

41 301.5 64.1 193 2 US-08-383-621-1 Sequence 1, Appli

42 301.5 64.1 193 3 US-08-459-906-1 Sequence 1, Appli

43 301.5 64.1 216 2 US-09-105-651-3 Sequence 3, Appli

44 299.5 63.7 190 1 US-07-963-331D-4 Sequence 1, Appli

45 299.5 63.7 191 1 US-07-942-523-1 Sequence 1, Appli

#### ALIGNMENTS

RESULT 1

US-08-093-483-1

Sequence 1, Application US/08093383

Patent No. 5489529

GENERAL INFORMATION:

APPLICANT: DeBoer, Herman A.

APPLICANT: Heyneker, Herbert L.

APPLICANT: Seeburg, Peter H.

TITLE OF INVENTION: DNA for Expression of Bovine Growth Hormone

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genentech, Inc.

STREET: 460 Point San Bruno Blvd

CITY: South San Francisco

STATE: California

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: patin (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/093,483

FILING DATE: 14-JUL-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/619827

FILING DATE: 28-NOV-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/198824

FILING DATE: 05-APR-1988

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 06/632361

FILING DATE: 19-JUL-1984

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 06/303687

FILING DATE: 18-SEP-1981

ATTORNEY/AGENT INFORMATION:

NAME: Johnston, Sean A.

REGISTRATION NUMBER: P35,910

REFERENCE/DOCKET NUMBER: 44C4

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/225-3562

TELEFAX: 415/952-9881

TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 192 amino acids

TYPE: amino acid

COMPUTER READABLE FORM:  
MEDIUM TYPE: ELECTRONIC DISK

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

US/06/383,621  
APPLICATION NUMBER:  
FILING DATE: 06-FEB-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/766,142  
FILING DATE: 25-SEP-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Tsevdos, Estelle J.

REFERENCE/DOI REL. NUMBER: 41,278-0.  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 203-421-2756  
TELEFAX: 203-421-2971  
TELEX: 203-710-474-4059  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 194 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
3-383-621-4

04 PSNKEEQKSNLEJRIKISLELLIQSWUEPVO 94

Sequence 4, Application US/06459906  
 Patent No. 6010999  
 GENERAL INFORMATION:

APPLICANT: Buckwalter, Brian L.  
APPLICANT: Cady, Susan M.  
APPLICANT: Shieh, Hong-Ming

TITLE OF INVENTION: Stabilization of Somatotropins and other  
 TITLE OF INVENTION: proteins by Modification of Cysteine Residues  
 NUMBER OF SEQUENCES: 11

ADDRESSEE: American Cyanamid Company  
STREET: One Cyanamid Plaza  
CITY: Wayne

Account: 0-0-0-0  
ZIP: 07470-8426  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

ATTORNEY/AGENT INFORMATION:  
CLASSIFICATION: 514  
FILING DATE: 02-JUN-1995  
NOTIFICATION NUMBER: 187867433200

NAME: \_\_\_\_\_

REGISTRATION NUMBER: 34,276  
 REFERENCE/DOCKET NUMBER: 31,278-03  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 201-831-3247  
 TELEFAX: 201-831-3305  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 194 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-459-906-4

Query Match 97.9% Score 460; DB 3; Length 194;

Best Local Similarity 98.9% Pred. No. 6,1e-51;  
 Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNAMLRAHRLHOLAFDTYOEFEAYIPKEOKYSFLONPQTSLSFSSEIPT 61  
 DB 4 FPTPLSRFDNAMLRAHRLHOLAFDTYOEFEAYIPKEOKYSFLONPQTSLSFSSEIPT 63  
 QY 62 PSNRETOOKSNLELLRISLLLSQSWLEPVO 92  
 DB 64 PSNRETOOKSNLELLRISLLLSQSWLEPVO 94

RESULT 5

US-08-589-028-10

Sequence 10, Application US/08589028

Patent No. 6087129

GENERAL INFORMATION:

APPLICANT: Newgard, Christopher B.

APPLICANT: Halban, Philippe A.

APPLICANT: No. 6087129mington, Karl D.

APPLICANT: Clark, Samuel A.

APPLICANT: Thigpen, Anice E.

APPLICANT: Quaade, Christian

APPLICANT: Kruse, Fred

TITLE OF INVENTION: Recombinant Expression of Proteins From

TITLE OF INVENTION: Secretory Cell Lino. S

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P. O. Box 4431

CITY: Houston

STATE: TX

COUNTRY: USA

ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/589,028

FILING DATE: Concurrently Herewith.

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Highlander, Steven L.

REGISTRATION NUMBER: 47,642

REFERENCE/DOCKET NUMBER: US/08/456\HY

TELEPHONE: (512) 418-3000

TELEFAX: (512) 474-7577

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 217 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

US-08-589-028-10

Query Match

97.9% Score 460; DB 3; Length 217;

Best Local Similarity 98.9% Pred. No. 7,1e-51;  
 Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNAMLRAHRLHOLAFDTYOEFEAYIPKEOKYSFLONPQTSLSFSSEIPT 61  
 DB 27 FPTPLSRFDNAMLRAHRLHOLAFDTYOEFEAYIPKEOKYSFLONPQTSLSFSSEIPT 86

QY 62 PSNRETOOKSNLELLRISLLLSQSWLEPVO 92

DB 87 PSNRETOOKSNLELLRISLLLSQSWLEPVO 117

RESULT 6

US-08-784-582-10

Sequence 10, Application US/08784582

Patent No. 6116707

GENERAL INFORMATION:

APPLICANT: Newgard, Christopher B.

APPLICANT: Halban, Philippe A.

APPLICANT: No. 6116707mington, Karl D.

APPLICANT: Clark, Samuel A.

APPLICANT: Thigpen, Anice E.

APPLICANT: Quaade, Christian

APPLICANT: Kruse, Fred

APPLICANT: McGarry, Dennis

TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM

TITLE OF INVENTION: SECRETORY CELL LINES

NUMBER OF SEQUENCES: 79

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P. O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/784,582

FILING DATE: Concurrently Herewith

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/328,427

FILING DATE: 15-OCT-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/589,028

FILING DATE: 19-JAN-1996

ATTORNEY/AGENT INFORMATION:

NAME: Highlander, Steven L.

REGISTRATION NUMBER: 37,642

REFERENCE/DOCKET NUMBER: US/08-514

TELECOMMUNICATION INFORMATION:

TELEPHONE: 512/418-3000

TELEFAX: 512/474-7577

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 217 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

US-08-784-582-10

Query Match 97.9% Score 460; DB 3; Length 217;

Best Local Similarity 98.9% Pred. No. 7,1e-51;

Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTPLSRFDNAMLRAHRLHOLAFDTYOEFEAYIPKEOKYSFLONPQTSLSFSSEIPT 61

DB 27 FPTPLSRFDNAMLRAHRLHOLAFDTYOEFEAYIPKEOKYSFLONPQTSLSFSSEIPT 86

OY 62 PSNRETOQKSNLELLRLISLLLIQSWLEPVQ 92  
 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!  
 DB 87 PSNRETOQKSNLELLRLISLLLIQSWLEPVQ 117

## RESULT 7

US-08-785-271-10  
 : Sequence 10, Application US/08785271  
 : Patent No. 6,941,766  
 : GENERAL INFORMATION:  
 : APPLICANT: Newgard, Christopher B.  
 : APPLICANT: Halban, Philippe A.  
 : APPLICANT: No. 6194176minution, Karl D.  
 : APPLICANT: Clark, Samuel A.  
 : APPLICANT: Thigpen, Anice E.  
 : APPLICANT: Ouade, Christian  
 : APPLICANT: Kruse, Fred  
 : TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM  
 : TITLE OF INVENTION: SECRETORY CELLS LINES  
 : NUMBER OF SEQUENCES: 56  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Almad, White & Decker  
 : STREET: P.O. Box 4433  
 : CITY: Houston  
 : STATE: Texas  
 : COUNTRY: USA  
 : ZIP: 77210

: COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.30  
 : APPLICATION NUMBER: US/08/785.271  
 : FILING DATE: Concurrently herewith  
 : CLASSIFICATION: 435  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 08/589,028  
 : FILING DATE: 19-JAN-1996

: ATTORNEY/AGENT INFORMATION:  
 : NAME: Highlander, Steven L.  
 : REGISTRATION NUMBER: 37,642  
 : REFERENCE/DOCKET NUMBER: UFSO-513  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 512/418-3000  
 : TELEFAX: 512/474-7577  
 : INFORMATION FOR SEQ ID NO: 10:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 217 amino acids  
 : TYPE: amino acid  
 : STRANDEDNESS: linear  
 : TOPOLOGY: linear

US-08-785-271-10

Query Match 97.9% Score 460; DB 3; Length 217;  
 Best Local Similarity 98.9% Pred. No. 7 1e-51;  
 Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 FPTIPLSRFDNMLRAHRLHQLAFDTYQFEFFAYIPKEOKYSFLQNPQTSLSFSSEIPT 61  
 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!  
 DB 27 FPTIPLSRFDNMLRAHRLHQLAFDTYQFEFFAYIPKEOKYSFLQNPQTSLSFSSEIPT 86  
 OY 62 PSNRETOQKSNLELLRLISLLLIQSWLEPVQ 92  
 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!  
 DB 87 PSNRETOQKSNLELLRLISLLLIQSWLEPVQ 117

## RESULT 8

US-08-759-628-11  
 : Sequence 11, Application US/08759628  
 : Patent No. 6225446  
 : GENERAL INFORMATION:  
 : APPLICANT: Altmann, Scott W.

: APPLICANT: Rock, Fernando L.  
 : APPLICANT: Bazan, J. Fernando  
 : APPLICANT: Kastelein, Robert A.  
 : TITLE OF INVENTION: MUTATIONAL VARIANTS OF MAMMALIAN PROTEINS  
 : NUMBER OF SEQUENCES: 11  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: DNAX Research Institute  
 : STREET: 901 California Avenue  
 : CITY: Palo Alto  
 : STATE: California  
 : COUNTRY: USA  
 : ZIP: 94304-1104

: COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/759,628  
 : FILING DATE: 05-DEC-1996  
 : CLASSIFICATION: 435  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 66/008,574  
 : FILING DATE: 06-DEC-1995

: ATTORNEY/AGENT INFORMATION:  
 : NAME: Ching, Edwin P.  
 : REGISTRATION NUMBER: 34,009  
 : REFERENCE/DOCKET NUMBER: 1X05520  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 415-852-9196  
 : TELEFAX: 415-496-1400

: INFORMATION FOR SEQ ID NO: 11:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 217 amino acids  
 : TYPE: amino acid  
 : STRANDEDNESS: single  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: protein

: FEATURE:  
 : NAME/KEY: Peptide  
 : LOCATION: 32..53  
 : FEATURE:  
 : NAME/KEY: Peptide  
 : LOCATION: 94..115  
 : FEATURE:  
 : NAME/KEY: Peptide  
 : LOCATION: 133..153

: NAME/KEY: Peptide  
 : LOCATION: 192..210  
 : OTHER INFORMATION: /note: "The peptides above are  
 : OTHER INFORMATION: depicted in Figure 1"  
 : US-08-759-628-11

Query Match 97.9% Score 460; DB 3; Length 217;  
 Best Local Similarity 98.9% Pred. No. 7 1e-51;  
 Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 FPTIPLSRFDNMLRAHRLHQLAFDTYQFEFFAYIPKEOKYSFLQNPQTSLSFSSEIPT 61  
 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!  
 DB 27 FPTIPLSRFDNMLRAHRLHQLAFDTYQFEFFAYIPKEOKYSFLQNPQTSLSFSSEIPT 86  
 OY 62 PSNRETOQKSNLELLRLISLLLIQSWLEPVQ 92  
 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!  
 DB 87 PSNRETOQKSNLELLRLISLLLIQSWLEPVQ 117

## RESULT 9

US-09-284-878-1  
 : Sequence 1, Application US/09284878  
 : Patent No. 6342375  
 : GENERAL INFORMATION:  
 : APPLICANT: Olazaran, Martha Guerrero



APPLICANT: Saldana, Hugo Barrera  
APPLICANT: Salvado, Jose Maria Viader  
TITLE OF INVENTION: Genetically Modified Methylophilic P. pastoris Yeast for the  
TITLE OF INVENTION: Production and Secretion of the Human Growth Hormone  
FILE REFERENCE: 1829.0010000  
CURRENT APPLICATION NUMBER: US/09/284,878  
PRIOR FILING DATE: 1999-07-21  
PRIOR APPLICATION NUMBER: PCT/MX97/00033  
PRIOR FILING DATE: 1997-10-24  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 217  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-284,878-1

Query Match: 97.9%; Score 460; DB 4; Length 241;  
Best Local Similarity 98.9%; Pred. No. 8.2e-51;  
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTIPSLRFDNAMLRAHRLHQLAFDTYQFEAYIPKEQKYSFLONPOTSLSFSES IPT 61  
|||||  
Db 27 FPTIPSLRFDNAMLRAHRLHQLAFDTYQFEAYIPKEQKYSFLONPOTSLSFSES IPT 86  
|||||

QY 62 PSNRRETOOKSNLELLRLISLLIQSWLEPVQ 92  
|||||  
Db 87 PSNRRETOOKSNLELLRLISLLIQSWLEPVQ 117  
|||||

RESULT 10  
US-09-424,620B-25  
Sequence 25, Application US/09/24620B  
Patent No. 6391585  
GENERAL INFORMATION:  
APPLICANT: HANIL SYNTHETIC FIBER CO., LTD.  
JANG, Ki-Ryong  
MOON, Jae-Woong  
HAE, Cheon-Soon  
YANG, Doo-Suk  
LEE, Jee-Won  
SEONG, Baik-Lin  
TITLE OF INVENTION: Process for preparing recombinant proteins using highly  
efficient expression vector from Saccharomyces cerevisiae  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BACHMAN & LAPORTE, P.C.  
STREET: Suite 1201, 900 Chapel Street  
CITY: New Haven  
STATE: Connecticut  
COUNTRY: U.S.A.  
ZIP: 06510-2802  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM  
OPERATING SYSTEM: WINDOWS 95/98  
SOFTWARE: MS WORD  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/424,620B  
FILING DATE: 24-Nov-03 6391585 1499  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 241 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: PROTEIN  
SEQUENCE DESCRIPTION: SEQ ID NO: 25:  
US-09-424-620B-25

Query Match: 97.9%; Score 460; DB 4; Length 241;  
Best Local Similarity 98.9%; Pred. No. 8.2e-51;  
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTIPSLRFDNAMLRAHRLHQLAFDTYQFEAYIPKEQKYSFLONPOTSLSFSES IPT 61  
|||||  
Db 51 FPTIPSLRFDNAMLRAHRLHQLAFDTYQFEAYIPKEQKYSFLONPOTSLSFSES IPT 110  
|||||

QY 62 PSNRRETOOKSNLELLRLISLLIQSWLEPVQ 92  
|||||  
Db 111 PSNRRETOOKSNLELLRLISLLIQSWLEPVQ 141  
|||||

RESULT 11  
US-09-280-030-66  
Sequence 66, Application US/09/280030A  
Patent No. 6506595  
GENERAL INFORMATION:  
APPLICANT: Sato, Seiji  
APPLICANT: Higashikami, Naohiko  
APPLICANT: Kudo, Toshiyuki  
TITLE OF INVENTION: DNAs encoding new fusion proteins and processes for  
TITLE OF INVENTION: PREPARING USEFUL POLYPEPTIDES THROUGH EXPRESSION OF THE  
TITLE OF INVENTION: DNAs  
FILE REFERENCE: 382.1026  
CURRENT APPLICATION NUMBER: US/09/280.030A  
CURRENT FILING DATE: 1999-03-26  
EARLIER APPLICATION NUMBER: JP10-87339/1998  
EARLIER FILING DATE: 1998-03-31  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 66  
LENGTH: 245  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Designated is  
OTHER INFORMATION: an amino acid sequence of MMDsp-MWmp20 TEV-G-GH  
US-09-280-030-66

Query Match: 97.9%; Score 460; DB 4; Length 245;  
Best Local Similarity 98.9%; Pred. No. 8.4e-51;  
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTIPSLRFDNAMLRAHRLHQLAFDTYQFEAYIPKEQKYSFLONPOTSLSFSES IPT 61  
|||||  
Db 55 FPTIPSLRFDNAMLRAHRLHQLAFDTYQFEAYIPKEQKYSFLONPOTSLSFSES IPT 114  
|||||

QY 62 PSNRRETOOKSNLELLRLISLLIQSWLEPVQ 92  
|||||  
Db 115 PSNRRETOOKSNLELLRLISLLIQSWLEPVQ 145  
|||||

RESULT 12  
US-08-784-592-71  
Sequence 71, Application US/08/784592  
Patent No. 6110707  
GENERAL INFORMATION:  
APPLICANT: Newgard, Christopher B.  
APPLICANT: Halban, Philippe A.  
APPLICANT: No. 6110707minington, Karl D.  
APPLICANT: Clark, Samuel A.  
APPLICANT: Thigpen, Anice E.  
APPLICANT: Ouade, Christian  
APPLICANT: Kruse, Fred  
APPLICANT: McCarry, Dennis  
TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM  
TITLE OF INVENTION: SECRETORY CELL LINES  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Ver. on #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/784,582
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,427
FILING DATE: 15-OCT-1996
APPLICATION NUMBER: US 08/589,028
FILING DATE: 19-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: JTSO-514
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 71:
SEQUENCE CHARACTERISTICS:
LENGTH: 274 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-784-582-71

Query Match          97.9%   Score 460: DB 3: Length 274:
Best Local Similarity 98.9%   Pred. No. 9.9e 51:
Matches 90: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

QY  2 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEFFAYIPKEQKYSFLQNPQTSLSFSSEIPT 61
DB  27 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEFFAYIPKEQKYSFLQNPQTSLSFSSEIPT 86

QY  62 PSNREETOQKSNLELLRISILLIQSWLEPVQ 92
DB  87 PSNREETOQKSNLELLRISILLIQSWLEPVQ 117

RESULT 14
US-09-465-461-1
Sequence 1: Application US/09/465/461
Patent No. 6348444
GENERAL INFORMATION:
APPLICANT: CHAPPEL, Scott
TITLE OF INVENTION: Human Growth Hormone to stimulate hematopoiesis and immune re
FILE REFERENCE: after hematopoietic stem cell transplantation in humans
CURRENT APPLICATION NUMBER: US/09/465,461
PRIOR FILING DATE: 1999-12-17
PRIOR APPLICATION NUMBER: 60/112,668
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 191
TYPE: PR:
ORGANISM: homo sapiens
US 09-465-461-1

Query Match          96.6%   Score 454: DB 4: Length 191:
Best Local Similarity 97.8%   Pred. No. 3.5e-50:
Matches 85: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

QY  2 FPTPLSRFLFNAMLRHRLHQLAFDTYQFEFFAYIPKEQKYSFLQNPQTSLSFSSEIPT 61
DB  1 FPTPLSRFLFNAMLRHRLHQLAFDTYQFEFFAYIPKEQKYSFLQNPQTSLSFSSEIPT 60

QY  62 PSNREETOQKSNLELLRISILLIQSWLEPVQ 92
DB  61 PSNREETOQKSNLELLRISILLIQSWLEPVQ 91

RESULT 15
US-08-187-756C-4
Sequence 4: Application US/08187756C

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 12:04:45 : Search time 38.5606 seconds  
(without alignments)  
347.945 Million cell updates/sec

Title: US-09-423-100-2

Perfect score: 470

Sequence: 1 MFPTPLSRFLDNAMLAHRR.....NLEELRISILIIOSWLRPVQ 42

Scoring table: HIGS0M62

Gapop 10.0, Gapext 0.5

Searched: 54,946 seqs, 1459,7476 residues

Total number of hits satisfying chosen parameters: 44

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA:

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:  
2: /cgn2\_6/ptodata/1/pubpaa/PC1\_NEW\_PUB.pep:  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:  
6: /cgn2\_6/ptodata/1/pubpaa/PC1US\_PUBCOMB.pep:  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	470	100.0	92	14	US-10-054-83-2 Sequence 2, Appl
2	470	100.0	150	14	US-10-054-83-7 Sequence 7, Appl
3	460	97.9	191	11	US-09-984-010-23 Sequence 23, Appl
4	460	97.9	191	12	US-10-153-207-1 Sequence 1, Appl
5	460	97.9	191	12	US-10-400-177-1 Sequence 1, Appl
6	460	97.9	191	12	US-10-400-708-1 Sequence 1, Appl
7	460	97.9	214	12	US-10-153-207-6 Sequence 6, Appl
8	460	97.9	217	9	US-09-929-918-9 Sequence 9, Appl
9	460	97.9	245	9	US-09-280-010-66 Sequence 66, Appl
10	454	96.6	217	9	US-09-853-688-2 Sequence 2, Appl
11	454	96.6	217	12	US-09-969-748C-4 Sequence 4, Appl
12	453	96.4	217	10	US-09-804-409A-16 Sequence 16, Appl
13	447	95.1	217	9	US-09-853-688-4 Sequence 4, Appl
14	444	94.5	191	12	US-09-824-200-12 Sequence 12, Appl
15	396	84.3	217	9	US-09-850-887-3 Sequence 3, Appl

16	381	81.1	163	15	US-10-043-487-350	Sequence 350, App
17	381	81.1	191	12	US-10-153-207-2	Sequence 2, Appl
18	373	79.4	229	15	US-10-103-313-411	Sequence 411, App
19	345	73.4	246	15	US-10-188-246-18	Sequence 18, Appl
20	302.5	64.1	190	12	US-10-153-207-3	Sequence 3, Appl
21	260	55.3	49	14	US-10-054-873-1	Sequence 1, Appl
22	260	55.3	107	14	US-10-054-873-6	Sequence 6, Appl
23	174.5	37.1	188	15	US-10-191-879-19	Sequence 19, Appl
24	174.5	37.1	210	15	US-10-191-879-10	Sequence 10, Appl
25	162	34.5	136	15	US-10-043-487-337	Sequence 337, App
26	161.5	34.4	386	10	US-09-887-569A-2	Sequence 2, Appl
27	151.5	34.4	388	12	US-10-122-746-4	Sequence 4, Appl
28	142	30.2	54	9	US-09-876-478-14	Sequence 14, Appl
29	131	27.9	198	15	US-10-140-293-32	Sequence 32, Appl
30	126	26.8	199	15	US-10-140-293-33	Sequence 33, Appl
31	125	26.6	199	15	US-10-140-293-26	Sequence 26, Appl
32	125	26.6	199	15	US-10-140-293-27	Sequence 27, Appl
33	124	25.4	199	15	US-10-140-293-31	Sequence 31, Appl
34	123	26.2	199	15	US-10-140-293-22	Sequence 22, Appl
35	122	26.0	199	15	US-10-140-293-21	Sequence 21, Appl
36	121.5	25.9	124	14	US-10-043-869-26	Sequence 26, Appl
37	121.5	25.9	208	15	US-10-191-879-2	Sequence 2, Appl
38	121.5	25.9	253	14	US-10-036-869-27	Sequence 27, Appl
39	119	25.3	199	15	US-10-140-293-24	Sequence 24, Appl
40	118	25.1	199	15	US-10-140-293-18	Sequence 18, Appl
41	118	25.1	199	15	US-10-140-293-19	Sequence 19, Appl
42	117	24.9	197	15	US-10-140-293-15	Sequence 15, Appl
43	116	24.7	197	15	US-10-140-293-13	Sequence 13, Appl
44	116	24.7	199	12	US-10-153-207-4	Sequence 4, Appl
45	116	24.7	199	15	US-10-140-293-20	Sequence 20, Appl

#### ALIGNMENTS

#### RESULT 1

US-10-054-873-2

: Sequence 2, Application US/10054873

: Publication No. US20020164712A1

: GENERAL INFORMATION:

: APPLICANT: Gan, Zhong Ru

: TITLE OF INVENTION: Chimeric Protein Containing an Intramolecular Chaperone-like Sequence

: NUMBER OF SEQUENCES: 7

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Townsend and Townsend and Crew LLP

: STREET: Two Embarcadero Center, Eighth Floor

: CITY: San Francisco

: STATE: California

: COUNTRY: USA

: ZIP: 94111-4934

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: PatentIn Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/10/054-873

: FILING DATE: 22-Jan-2002

: CLASSIFICATION: <Unknown>

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: WO PCT/CN98/00052

: FILING DATE: 31-MAR-1998

: APPLICATION NUMBER: US 09/423,100

: FILING DATE: 11-DEC-2000

: ATTORNEY/AGENT INFORMATION:

: NAME: MYCtoft, Frank J

: REGISTRATION NUMBER: 46,946

: REFERENCE/DOCKET NUMBER: 020167-000130US

: INFORMATION FOR SEQ ID NO: 2:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 92 amino acids

: TYPE: amino acid



```
; PRIOR FILING DATE: 1992-10-13
; PRIOR APPLICATION NUMBER: 07/875,204
; PRIOR FILING DATE: 1992-04-27
; PRIOR APPLICATION NUMBER: 07/428,066
; PRIOR FILING DATE: 1989-10-26
; PRIOR APPLICATION NUMBER: 07/264,611
; PRIOR FILING DATE: 1988-10-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-400-377-1
```

```
Query Match          97.9%   Score 460; DB 12; Length 191;
Best Local Similarity 98.9%   Pred. No. 780-45;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTPLSKLFONAMLAHRLHQLAFDYOFEEAYIPKEQKYSFLONPQTSLSSESPT 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 FPTPLSKLFONAMLAHRLHQLAFDYOFEEAYIPKEQKYSFLONPQTSLSSESPT 60

QY 62 PSNREETOOKSNLELLRLISLLLIQSWLEPVQ 92
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 PSNREETOOKSNLELLRLISLLLIQSWLEPVQ 91
```

```
RESULT 5
US-10-400-377-1
; Sequence 1, Application US/10400377
; Publication No. US20030162949A1
; GENERAL INFORMATION:
; APPLICANT: COX III, George N
; APPLICANT: Bolder Biotechnology, Inc.
; TITLE OF INVENTION: Derivatives of Growth Hormone and Related Proteins
; FILE REFERENCE: 4152-1-PUS
; CURRENT APPLICATION NUMBER: US/10/400,377
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US/09/462,941
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: 60/052,516
; PRIOR FILING DATE: 1997-07-14
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-400-377-1
```

```
Query Match          97.9%   Score 460; DB 12; Length 191;
Best Local Similarity 98.9%   Pred. No. 780-45;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTPLSKLFONAMLAHRLHQLAFDYOFEEAYIPKEQKYSFLONPQTSLSSESPT 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 FPTPLSKLFONAMLAHRLHQLAFDYOFEEAYIPKEQKYSFLONPQTSLSSESPT 60

QY 62 PSNREETOOKSNLELLRLISLLLIQSWLEPVQ 92
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 PSNREETOOKSNLELLRLISLLLIQSWLEPVQ 91
```

```
RESULT 6
US-10-400-798-1
; Sequence 1, Application US/10400708
; Publication No. US20030166865A1
; GENERAL INFORMATION:
; APPLICANT: COX III, George N
; APPLICANT: Bolder Biotechnology, Inc.
; TITLE OF INVENTION: Derivatives of Growth Hormone and Related Proteins
; FILE REFERENCE: 4152-1-PUS
```

```
; CURRENT APPLICATION NUMBER: US/10/400,798
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US/09/462,941
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: 60/052,516
; PRIOR FILING DATE: 1997-07-14
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-400-798-1
```

```
Query Match          97.9%   Score 460; DB 12; Length 191;
Best Local Similarity 98.9%   Pred. No. 780-45;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTPLSKLFONAMLAHRLHQLAFDYOFEEAYIPKEQKYSFLONPQTSLSSESPT 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 FPTPLSKLFONAMLAHRLHQLAFDYOFEEAYIPKEQKYSFLONPQTSLSSESPT 60

QY 62 PSNREETOOKSNLELLRLISLLLIQSWLEPVQ 92
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 PSNREETOOKSNLELLRLISLLLIQSWLEPVQ 91
```

```
RESULT 7
US-10-153-207-6
; Sequence 6, Application US/10153207
; Publication No. US20030153003A1
; GENERAL INFORMATION:
; APPLICANT: James A. Wells
; APPLICANT: Brian C. Cunningham
; TITLE OF INVENTION: GROWTH HORMONE VARIANTS
; FILE REFERENCE: 669,12-US-C7
; CURRENT APPLICATION NUMBER: US/10/153,207
; CURRENT FILING DATE: 2002-05-22
; PRIOR APPLICATION NUMBER: 08/479,884
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/190,723
; PRIOR FILING DATE: 1994-02-02
; PRIOR APPLICATION NUMBER: 07/960,227
; PRIOR FILING DATE: 1992-10-13
; PRIOR APPLICATION NUMBER: 07/875,204
; PRIOR FILING DATE: 1992-04-27
; PRIOR APPLICATION NUMBER: 07/428,066
; PRIOR FILING DATE: 1989-10-26
; PRIOR APPLICATION NUMBER: 07/264,611
; PRIOR FILING DATE: 1988-10-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-153-207-6
```

```
Query Match          97.9%   Score 460; DB 12; Length 214;
Best Local Similarity 98.9%   Pred. No. 90-45;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTPLSKLFONAMLAHRLHQLAFDYOFEEAYIPKEQKYSFLONPQTSLSSESPT 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24 FPTPLSKLFONAMLAHRLHQLAFDYOFEEAYIPKEQKYSFLONPQTSLSSESPT 83

QY 62 PSNREETOOKSNLELLRLISLLLIQSWLEPVQ 92
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 84 PSNREETOOKSNLELLRLISLLLIQSWLEPVQ 114
```

```
RESULT 8
US-09-929-918-9
```

; Sequence 9, Application US/09929918  
; Patent No. US20020090678A1  
; GENERAL INFORMATION:  
; APPLICANT: Kordyum, Vitaliy A.  
; APPLICANT: Chernykh, Svitlana I.  
; APPLICANT: Slavchenko, Iryna Yu.  
; APPLICANT: Vozlianov, Oleksandr  
; TITLE OF INVENTION: PHAGE-DEPENDENT SUPER PRODUCTION OF  
; TITLE OF INVENTION: BIOLOGICALLY ACTIVE PROTEIN AND PEPTIDES  
; FILE REFERENCE: PHAGE.008A  
; CURRENT APPLICATION NUMBER: US/09/929,918  
; CURRENT FILING DATE: 2001-08-15  
; PRIOR APPLICATION NUMBER: 09/318,288  
; PRIOR FILING DATE: 1999-05-25  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 217  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-929-918 9

Query Match 97.9% Score 460; DB 9; Length 217;  
Best Local Similarity 98.9% Pred. No. 9.2e-45;  
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTPLSKLFNMLRAHRLHQLAFDYQFEFEAYIPKEOKYSLQNPOTSLCFSESPT 61  
|||||  
DB 27 FPTPLSKLFNMLRAHRLHQLAFDYQFEFEAYIPKEOKYSLQNPOTSLCFSESPT 86  
|||||  
QY 62 PSNREETOQKSNLELLRLISLLIQSWLEPVO 12  
|||||  
DB 87 PSNREETOQKSNLELLRLISLLIQSWLEPVO 117  
|||||

RESULT 9  
US-09-280-010-66  
; Sequence 66, Application US/0928030A  
; Patent No. US20010021515A1  
; GENERAL INFORMATION:  
; APPLICANT: Sato, Seiji  
; APPLICANT: Higashikuni, Naohiko  
; APPLICANT: Kudo, Toshiyuki  
; APPLICANT: Kondo, Masaaki  
; TITLE OF INVENTION: DNAS ENCODING NEW FUSION PROTEINS AND PROCESSES FOR  
; TITLE OF INVENTION: PREPARING USEFUL POLYPEPTIDES THROUGH EXPRESSION OF THE  
; TITLE OF INVENTION: DNAS  
; FILE REFERENCE: 382,1026  
; CURRENT APPLICATION NUMBER: US/09/280,010A  
; CURRENT FILING DATE: 1999-03-26  
; EARLIER APPLICATION NUMBER: JP10-87339/1998  
; EARLIER FILING DATE: 1998-03-31  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 66  
; LENGTH: 245  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Designated as  
; OTHER INFORMATION: an amino acid sequence of MWsp:MWmp20-THV-G-GH  
US-09-280-010-66

Query Match 97.9% Score 460; DB 9; Length 245;  
Best Local Similarity 98.9% Pred. No. 1.1e-44;  
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTPLSKLFNMLRAHRLHQLAFDYQFEFEAIPKEOKYSLQNPOTSLCFSESPT 61  
|||||  
DB 55 FPTPLSKLFNMLRAHRLHQLAFDYQFEFEAYIPKEOKYSLQNPOTSLCFSESPT 114  
|||||  
QY 62 PSNREETOQKSNLELLRLISLLIQSWLEPVO 92  
|||||

Db 115 PSNREETOQKSNLELLRLISLLIQSWLEPVO 145

## RESULT 10

US-09-853-688-2  
; Sequence 2, Application US/09853688  
; Patent No. US20020081605A1  
; GENERAL INFORMATION:  
; APPLICANT: COOPER, DAVID N.  
; APPLICANT: PROCTER, ANNIE M.  
; APPLICANT: GREGORY, JOHN  
; APPLICANT: MILLAR, DAVID S.  
; TITLE OF INVENTION: METHOD FOR DETECTING GROWTH HORMONE VARIATIONS IN  
; TITLE OF INVENTION: HUMANS, THE VARIATIONS AND THEIR USES  
; FILE REFERENCE: W0M78  
; CURRENT APPLICATION NUMBER: US/09/853,688  
; CURRENT FILING DATE: 2001-05-14  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 217  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-853-688-2

Query Match 96.6% Score 454; DB 9; Length 217;  
Best Local Similarity 97.8% Pred. No. 4.5e-44;  
Matches 89; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 FPTPLSKLFNMLRAHRLHQLAFDYQFEFEAYIPKEOKYSLQNPOTSLCFSESPT 61  
|||||  
DB 27 FPTPLSKLFNMLRAHRLHQLAFDYQFEFEAYIPKEOKYSLQNPOTSLCFSESPT 86  
|||||  
QY 62 PSNREETOQKSNLELLRLISLLIQSWLEPVO 92  
|||||  
DB 87 PSNREETOQKSNLELLRLISLLIQSWLEPVO 117  
|||||

## RESULT 11

US-09-969-748C-4  
; Sequence 4, Application US/09969748C  
; Publication No. US20030161809A1  
; GENERAL INFORMATION:  
; APPLICANT: ARIZEKE PHARMACEUTICALS, INC.  
; APPLICANT: HOUSTON, Lou, L.  
; APPLICANT: SHERIDAN, Philip, J.  
; APPLICANT: HAWLEY, Stephen  
; APPLICANT: GLYNN, Jacqueline, M.  
; APPLICANT: CHAPIN, Steven  
; APPLICANT: BASU, Amresh  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TRANSPORT OF BIOLOGICALLY AC  
; TITLE OF INVENTION: AGENTS ACROSS CELLULAR BARRIERS  
; FILE REFERENCE: 057226-0304  
; CURRENT APPLICATION NUMBER: US/09/969,748C  
; CURRENT FILING DATE: 2002-12-10  
; PRIOR APPLICATION NUMBER: US 63/267,501  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: US 60/248,819  
; PRIOR FILING DATE: 2000-11-14  
; PRIOR APPLICATION NUMBER: US 60/248,478  
; PRIOR FILING DATE: 2000-11-13  
; PRIOR APPLICATION NUMBER: US 60/217,929  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 115  
; SOFTWARE: Patent In version 3.0  
; SEQ ID NO 4  
; LENGTH: 217  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-969-748C-4

Query Match 96.6% Score 454; DB 12; Length 217;  
Best Local Similarity 97.8% Pred. No. 4.5e-44;

```

Matches 89; Conservative 0; Mismatch: 5 2; Indels 0; Gaps 0;
QY 2 FPTIPLSRFDNMLRAHRLHQLAFDTYQEFEEAYIPKQKYSFLONPOTSLSFSESPT 61
DB 27 FPTIPLSRFDNMLRAHRLHQLAFDTYQEFEEAYIPKQKYSFLONPOTSLSFSESPT 86
QY 62 PSNREETOOKSNLELRISLLLIQSWLEPVQ 92
DB 87 PSNREETOOKSNLELRISLLLIQSWLEPVQ 117

RESULT 12
US-09-804-409A-16
: Sequence 16, Application US/09804409A
: Patent No. US2002015100A1
: GENERAL INFORMATION:
: APPLICANT: KIEFER, TIMOTHY J.
: APPLICANT: CHENG, ANTHONY T.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATED PROTEIN
: FILE REFERENCE: 025996/027, 8721
: CURRENT APPLICATION NUMBER: US/09/804,409A
: CURRENT FILING DATE: 2001-03-12
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: Patent In Ver. 2.1
: SEQ ID NO: 16
: LENGTH: 217
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-804-409A-16

Query Match 96.4%; Score 453; DB 10; Length 217;
Best Local Similarity 97.8%; Pred. No. 5, 8e-43;
Matches 89; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 FPTIPLSRFDNMLRAHRLHQLAFDTYQEFEEAYIPKQKYSFLONPOTSLSFSESPT 61
DB 27 FPTIPLSRFDNMLRAHRLHQLAFDTYQEFEEAYIPKQKYSFLONPOTSLSFSESPT 86
QY 62 PSNREETOOKSNLELRISLLLIQSWLEPVQ 92
DB 87 PSNREETOOKSNLELRISLLLIQSWLEPVQ 117

RESULT 13
US-09-853-688-4
: Sequence 4, Application US/09853688
: Patent No. US20020081605A1
: GENERAL INFORMATION:
: APPLICANT: COOPER, DAVID N.
: APPLICANT: PROCTER, ANNIE M.
: APPLICANT: GREGORY, JOHN
: APPLICANT: MILLAR, DAVID S.
: TITLE OF INVENTION: METHOD FOR DETECTING GROWTH HORMONE VARIATIONS IN
: HUMANS, THE VARIATIONS, AND THEIR USES
: FILE REFERENCE: WCM78
: CURRENT APPLICATION NUMBER: US/09/853,688
: CURRENT FILING DATE: 2001-05-14
: NUMBER OF SEQ ID NOS: 56
: SOFTWARE: Patent In Ver. 2.1
: SEQ ID NO: 4
: LENGTH: 217
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-853-688-4

Query Match 95.1%; Score 447; DB 9; Length 217;
Best Local Similarity 96.7%; Pred. No. 1, 8e-43;
Matches 88; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 FPTIPLSRFDNMLRAHRLHQLAFDTYQEFEEAYIPKQKYSFLONPOTSLSFSESPT 61
DB 27 FPTIPLSRFDNMLRAHRLHQLAFDTYQEFEEAYIPKQKYSFLONPOTSLSFSESPT 86

```

```

QY 62 PSNREETOOKSNLELRISLLLIQSWLEPVQ 92
DB 87 PSNREETOOKSNLELRISLLLIQSWLEPVQ 117

RESULT 14
US-09-824-200-12
: Sequence 12, Application US/09824200
: Patent No. US20030167531A1
: GENERAL INFORMATION:
: APPLICANT: RUSSELL, DOUGLAS A.
: APPLICANT: SCHLITTLER, MICHAEL
: TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF BIOACTIVE, AUTHENTIC
: POLYPEPTIDES FROM PLANTS
: FILE REFERENCE: 16712,003
: CURRENT APPLICATION NUMBER: US/09/824,200
: PRIOR FILING DATE: 2001-04-03
: PRIOR APPLICATION NUMBER: 99/194,217
: PRIOR FILING DATE: 2000-04-04
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: Patent In Ver. 2.1
: SEQ ID NO: 12
: LENGTH: 191
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-824-200-12

Query Match 94.5%; Score 443; DB 12; Length 191;
Best Local Similarity 96.7%; Pred. No. 5, 3e-43;
Matches 88; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 FPTIPLSRFDNMLRAHRLHQLAFDTYQEFEEAYIPKQKYSFLONPOTSLSFSESPT 61
DB 1 FPTIPLSRFDNMLRAHRLHQLAFDTYQEFEEAYIPKQKYSFLONPOTSLSFSESPT 60
QY 62 PSNREETOOKSNLELRISLLLIQSWLEPVQ 92
DB 61 PSNREETOOKSNLELRISLLLIQSWLEPVQ 91

RESULT 15
US-09-850-887-3
: Sequence 3, Application US/09850887
: Patent No. US20020009778A1
: GENERAL INFORMATION:
: APPLICANT: Lal, Preeti
: APPLICANT: Corley, Neil C.
: APPLICANT: Gargano, Gina
: TITLE OF INVENTION: THYROID AND PITUITARY MEMBRANE PROTEIN
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
: STREET: 3174 PORTER DRIVE
: CITY: PALO ALTO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC DOS/MS-DOS
: SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/850,887
: FILING DATE: 07 May 2001
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 09/087,678
: FILING DATE: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: CERRONE, MICHAEL C.
: REGISTRATION NUMBER: 39,132

```



```

: REFERENCE/DOCKET NUMBER: PF-0535 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (650) 855-0555
: TELEFAX: (650) 845-4166
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 217 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: GenBank
: CLONE: g406987
: SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-842-867 3

Query Match: H4 38: Score 492.1876, Length 217.
Best local similarity: H4 68: Prob. 6.1e-62
Matches: 77% conservative 62 Mismatches 8: Gaps 0:

Oy 2 FPTIPLSRLEFDMALRAHRLRLGLAFITYGFFHAYIPKDYKYSFLQNPQTSLSFSSESIPT 61
Db 27 FPTIPLSWLENTAVERAHLHLKLAFTYPKIFEAYIPKDYKYSFLRNPQTSLSFSSESIPT 86

Oy 62 PSNKEETQCKSNLELLRLISILLIQSWLEPVQ 92
Db 67 PSNKEETQCKSNLELLRLISILLIQSWLEPVQ 117

Search completed: September 15, 2003, 12:23:45
Job time: 34.5806 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:56:45 ; Search time 15.1685 seconds  
(without alignments)  
583,284 Million cell updates/sec

Title: US-09-423-100-2

Perfect score: 470

Sequence: 1 MFPTPLSRFLONAMLRHR.....NEELIKRISILLIQSWLEPVQ 92

Scoring table: BLOSUM62  
Gapop 10 0 , Gapext 0 0

Searchset: 2800k seqs, 60158662 res, 1.0e

Total number of hits satisfying chosen parameters: 283408

Minimum DB seq length: 5

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	460	97.9	217	1	STHO
2	460	97.9	217	2	somatotropin 1 pre
3	422	89.8	217	1	STHOV2
4	422	89.8	217	2	somatotropin 2 pre
5	402	85.5	217	2	167411
6	397	84.5	217	2	167409
7	396	84.3	212	2	167408
8	376	84.3	217	2	153267
9	381	81.1	217	1	LCRUC
10	381	81.1	217	2	E32435
11	359.5	76.5	215	2	A26449
12	310.5	66.1	216	2	B49159
13	307.5	65.4	190	2	PN0140
14	304.5	64.8	216	1	STMS
15	302.5	64.4	190	1	STHO
16	302.5	64.4	216	1	STRT
17	302.5	64.4	216	2	S49483
18	301.5	64.1	190	2	JKG219
19	301.5	64.1	216	1	STPG
20	301.5	64.1	216	2	I46145
21	301.5	64.1	216	2	JC4632
22	299.5	63.7	216	2	A37782
23	297.5	63.3	190	1	A61584
24	295.5	62.9	190	2	JS0429
25	289.5	61.6	217	1	STBO
26	289.5	61.6	217	1	STSH
27	289.5	61.6	217	1	STGT
28	289.5	61.6	217	2	S32682
29	278.5	59.3	216	2	JC1514

30	275.5	58.6	216	2	A60509
31	268.5	57.1	191	2	A60625
32	261	55.5	216	2	S04929
33	257.5	54.8	190	2	S21750
34	247.5	52.7	190	2	A56816
35	238.5	50.7	215	2	151188
36	237.5	50.5	215	2	JS0037
37	234	49.8	199	2	B32435
38	233.5	49.7	195	2	151250
39	225.5	48.0	183	2	A60623
40	206	43.8	87	4	167761
41	174.5	37.1	209	2	JT0483
42	171	36.4	163	2	JN0387
43	165.5	35.2	190	2	JC5682
44	165.5	35.2	210	2	150763
45	165.5	35.2	210	2	S36351

ALIGNMENT

## RESULT 1

STHO

somatotropin 1 precursor [validated] - human

N:Alternate names: growth hormone 1; GH-N; pituitary somatotropin

C:Species: Homo sapiens (man)

C:Date: 24-Apr-1984 ;sequence revision 10-Feb-1995 ;text change 08-Dec-2000

C:Accession: A93731; A32435; A93694; A94247; A90651; A93397; A91778; A91764; A90217

R:DeNoto, F.M.; Moore, D.D.; Goodman, H.M.

Nucleic Acids Res. 9, 3719-3730, 1981

A:Title: Human growth hormone DNA sequence and mRNA structure: possible alternative

A:Reference number: A93731; MUID:82014939; PMID:6259091

A:Accession: A93731

A:Molecule type: DNA

A:Residues: 1-217 <DEN>

A:Cross-references: GB:V00520

A>Note: the 20K short form somatotropin lacks residues 58-72 (32-46 in the active h-

R:Chen, E.Y.; Liao, Y.C.; Smith, D.H.; Barrera-Saldana, H.A.; Gellinas, R.E.; Seebur

Genomics 4, 479-497, 1989

A:Title: The human growth hormone locus: nucleotide sequence, biology, and evolutio

A:Reference number: A32435; MUID:89307273; PMID:2744760

A:Accession: A32435

A:Molecule type: DNA

A:Residues: 1-217 <CHE>

A:Cross-references: GB:J03071; MUID:q183148; PIDN:AAA52549.i; PMID:q183149

R:Koskam, W.; Rougeon, F.

Nucleic Acids Res. 7, 305-320, 1979

A:Title: Molecular cloning and nucleotide sequence of the human growth hormone stru

A:Reference number: A93694; MUID:89034477; PMID:986281

A:Accession: A93694

A:Molecule type: mRNA

A:Residues: 1-217 <ROS>

A:Cross-references: GB:V00519

A>Note: 35-Pro was also found

R:Martial, J.A.; Hallewell, R.A.; Baxter, J.D.; Goodman, H.M.

Science 205, 602-607, 1979

A:Title: Human growth hormone: complementary DNA cloning and expression in bacteria

A:Reference number: A94247; MUID:79203293; PMID:777496

A:Accession: A94247

A:Molecule type: mRNA

A:Residues: 1-217 <MAR>

R:Li, C.H.; Dixon, J.S.; Liu, W.K.

Arch. Biochem. Biophys. 133, 70-9, 1969

A:Title: Human pituitary growth hormone. XIX. The primary structure of the hormone.

A:Reference number: A90048; MUID: 9289202; PMID:5810834

A:Contents: annotation

R:Li, C.H.; Dixon, J.S.

Arch. Biochem. Biophys. 146, 233-236, 1971

A:Title: Human pituitary growth hormone. XXXI. The primary structure of the hormon

A:Reference number: A90051; MUID:72143935; PMID:5144027

A:Accession: A90051

A:Molecule type: protein

A:Residues: 27-94;96-217 <LIC>  
 R:Niall, H.D.  
 Nature New Biol. 230, 90-91, 1971  
 A:Title: Revised primary structure for human growth hormone.  
 A:Reference number: A93397; MUID:71139765; PMID:5275046  
 A:Accession: A93397  
 A:Molecule type: protein  
 A:Residues: 27-51 <NIA>  
 R:Niall, H.D.; Hogan, M.L.; Sauer, R.; Rosenblum, I.Y.; Greenwood, F.C.  
 Proc. Natl. Acad. Sci. U.S.A. 68, 866-869, 1971  
 A:Title: Sequences of pituitary and placental lactogenic and growth hormones: evolution  
 A:Reference number: A93778; MUID:71153968; PMID:5275528  
 A:Accession: A93778  
 A:Molecule type: protein  
 A:Residues: 119-120;157-159 <NI2>  
 R:Niall, H.D.  
 in Prolactin and Carcinogenesis, Paper, Fourth Fetus Workshop, Proceedings, Kluwer Academic Publishers, Dordrecht, 1981  
 A:Title: The chemistry of the human lactogenic hormones.  
 A:Reference number: A94427  
 A:Contents: annotation; somatotropin revision  
 R:Bewley, T.A.; Dixon, J.S.; Li, C.H.  
 Int. J. Pept. Protein Res. 4, 281-287, 1972  
 A:Title: Sequence comparison of human pituitary growth hormone, human chorionic somatomedin C, and human placental lactogen.  
 A:Reference number: A91764; MUID:73092028; PMID:4675454  
 A:Accession: A91764  
 A:Molecule type: protein  
 A:Residues: 27-217 <HEW>  
 R:Lewis, J.J.; Homwald, L.F.; Lewis, L.J.  
 Biochem. Biophys. Res. Commun. 92, 511-516, 1980  
 A:Title: The 20,000-dalton variant of human growth hormone: location of the amino acid differences.  
 A:Reference number: A90217; MUID:80130196; PMID:7356479  
 A:Contents: somatotropin, 20K short variant  
 A:Accession: A90217  
 A:Molecule type: protein  
 A:Residues: 46-57;73-80 <LEW>  
 R:Chapman, G.E.; Rogers, K.M.; Brattain, L.; Bradshaw, R.A.; Bates, G.J.; Turner, C.; Calkins, J.  
 J. Biol. Chem. 256, 2395-2401, 1981  
 A:Title: The 20,000 molecular weight variant of human growth hormone. Preparation and sequence determination.  
 A:Reference number: A92311; MUID:81117361; PMID:7462247  
 A:Contents: somatotropin, 20K short variant  
 A:Accession: A92311  
 A:Molecule type: protein  
 A:Residues: 27-57;73-79 <CHA>  
 R:Singh, R.N.P.; Sravey, B.K.; Lewis, L.J.; Lewis, U.J.  
 J. Protein Chem. 2, 425-436, 1983  
 A:Title: Human growth hormone peptide 1-43: isolation from pituitary glands.  
 A:Reference number: A61466  
 A:Accession: A61466  
 A:Molecule type: protein  
 A:Residues: 27-69 <SIN>  
 R:Robson, V.M.J.; Rae, I.D.; NG, F.  
 Biol. Chem. Hoppe-Seyler 371, 423-431, 1990  
 A:Title: Identification of the aspartimide structure in a previously-reported peptide.  
 A:Reference number: S09685; MUID:90134745; PMID:2378679  
 A:Accession: S09685  
 A:Molecule type: protein  
 A:Residues: 27-34;17-36-47 <ROB>  
 R:De Vos, A.M.; Ullsch, M.; Kossiakoff, A.A.  
 Science 255, 306-312, 1992  
 A:Title: Human growth hormone and extracellular domain of its receptor: crystal structure and functional analysis.  
 A:Reference number: A41728; MUID:92196577; PMID:1549776  
 A:Contents: annotation; x-ray crystallography, 2.8 angstroms  
 R:Gray, G.L.; Baldridge, J.S.; McKeown, K.S.; Heyneker, H.L.; Chang, C.N.  
 Gene 34, 247-254, 1985  
 A:Title: Periplasmic production of correctly processed human growth hormone in Escherichia coli.  
 A:Reference number: I41126; MUID:86137393; PMID:3912261  
 A:Accession: I84549  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-26 <RES>  
 A:Cross-references: GB:M14398; NID:q183158; PIDN:AAA52554.1; PID:q183159

C:Comment: The gene for this hormone is transcribed only in somatotrophic cells of the pituitary gland.  
 C:Comment: About 90% of somatotropin is the 22K long form.

C:Genetics:  
 A:Gene: GDB:GH1  
 A:Cross-references: GDB:119982; OMIM:139250  
 A:Map position: 17q23.1-17q23.3  
 A:Introns: 4/1; 57/3; 97/3; 152/3  
 A:Superfamily: prolactin  
 C:Keywords: alternative splicing; hormone; pituitary  
 F:1-26/Domain: signal sequence #status predicted <SIG>  
 F:27-217/Product: somatotropin 1, long form #status experimental <SOL>  
 F:27-69/Product: growth hormone 5K peptide #status experimental <SKP>  
 F:27-57,73-217/Product: somatotropin 1, short form #status experimental <SOS>  
 F:79-151,208-215/Disulfide bonds: #status experimental

Query Match: 97.9% Score 460; DR 1; Length 217;  
 Best Local Similarity 48.44; Pred. No. 4,40-42;  
 Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 FPTIPLSRLEFDNAMLRAHRLHQLAFDTYQFFEAYIPKEQKYSFLQNPOTLSLSESISPT 61  
 DB 27 FPTIPLSRLEFDNAMLRAHRLHQLAFDTYQFFEAYIPKEQKYSFLQNPOTLSLSESISPT 86  
 QY 62 PSNREETOQKSNLELLRISLLLIQSMLPEVQ 92  
 DB 87 PSNREETOQKSNLELLRISLLLIQSMLPEVQ 117

## RESULT 2

167410  
 Somatotropin - rhesus macaque  
 N:Alternate names: growth hormone  
 C:Species: Macaca mulatta (rhesus macaque)  
 C:Date: 31-May-1996 #sequence revision 11-May-1996 #text change 16-Jul-1999  
 C:Accession: 167410; A05094  
 R:Gollos, T.G.; Durning, M.; Fisher, J.M.; Fowler, P.D.  
 Endocrinology 133, 1744-1752, 1993  
 A:Title: Cloning of four growth hormone/chorionic somatomotropin-related complementary DNAs from rhesus macaque.  
 A:Reference number: 153267; MUID:94008724; PMID:8404617  
 A:Accession: 167410  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-217 <RES>  
 A:Cross-references: GR:116556; NID:q293114; PIUN:AAA18842.1; PID:q293115  
 R:Li, C.H.; Chung, D.; Lahm, H.W.; Stein, S.  
 Arch. Biochem. Biophys. 245, 287-291, 1986  
 A:Title: The primary structure of monkey pituitary growth hormone.  
 A:Reference number: A05094; MUID:86129460; PMID:3080959  
 A:Accession: A05094  
 A:Molecule type: protein  
 A:Residues: 27-59;101-178;180-217 <LIC>  
 A:Note: the monkey species is not identified in the reference  
 R:Raben, M.S.  
 Science 125, 883-884, 1957  
 A:Title: Preparation of growth hormone from pituitaries of man and monkey.  
 A:Reference number: A4774  
 A:Contents: annotation; identification of source organism  
 C:Superfamily: prolactin

## Query Match:

Best Local Similarity 48.44; Pred. No. 4,40-42;  
 Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 FPTIPLSRLEFDNAMLRAHRLHQLAFDTYQFFEAYIPKEQKYSFLQNPOTLSLSESISPT 61  
 DB 27 FPTIPLSRLEFDNAMLRAHRLHQLAFDTYQFFEAYIPKEQKYSFLQNPOTLSLSESISPT 86  
 QY 62 PSNREETOQKSNLELLRISLLLIQSMLPEVQ 92  
 DB 87 PSNREETOQKSNLELLRISLLLIQSMLPEVQ 117  
 RESULT 3

STHUV  
 somatotropin 2 precursor - human  
 N:Alternate names: growth hormone 2; growth hormone variant; hGH-V; placental somatotropin  
 N:Contains: somatotropin 2, long splice form; somatotropin 2, short splice form  
 C:Species: Homo sapiens (man)  
 C>Date: 17-Dec-1982 #sequence\_revision 10-Feb-1995 #text\_change 21-Jul-2000  
 C:Accession: D32435; B28072; A01511; I52104; A60711  
 R:Chen, E.Y.; Liao, Y.C.; Smith, D.H.; Barrera-Saldana, H.A.; Gelinas, K.E.; Seeburg, P.  
 Genomics 4: 479-497, 1989  
 A:Title: The human growth hormone locus: nucleotide sequence, biology, and evolution.  
 A:Reference number: A32435; MUID:89307277; PMID:2744760  
 A:Accession: D32435  
 A:Molecule type: DNA  
 A:Residues: 1-217 <CHE>  
 A:Cross-references: GB:J03071; NID:q183148; PIDN:AAA52552.1; PID:q183152  
 R:Cooke, N.E.; Kay, J.; Emery, J.G.; Liebhauer, S.A.  
 J. Biol. Chem. 263: 9001-9006, 1988  
 A:Title: Two distinct species of human growth hormone variant mRNA in the human placenta  
 A:Reference number: A92725; MUID:88243759; PMID:3379057  
 A:Accession: B28072  
 A:Molecule type: mRNA  
 A:Residues: 1-217 <COO>  
 A:Note: an alternative splice junction for intron 4 is used  
 C:Genetics:  
 A:Gene: GH2; GH2  
 A:Cross-references: GDB:119943; OMIM:139240  
 A:Map position: 17q22-17q24  
 A:Introns: 4/1; 57/3; 97/3; 152/3  
 C:Superfamily: prolactin  
 C:Keywords: alternative splicing; hormone; placenta  
 F:1-26/domain: signal sequence #status predicted <SIG>  
 F:27-256/product: somatotropin 2 splice form 2 #status predicted <MAI>  
 Query Match 89.8%; Score 422; DB 1; Length 256;  
 Best Local Similarity 92.3%; Pred. No. 6,7e-38;  
 Matches 84; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 A:Accession: A01511  
 A:Molecule type: DNA  
 A:Residues: 1-34; p.136-217 <SEE>  
 R:Igout, A.; Scippo, M.L.; Frankenne, F.; Henneen, G.  
 Arch. Int. Physiol. Biochim. 96: 63-67, 1988  
 A:Title: Cloning and nucleotide sequence of placental hGH-V cDNA.  
 A:Reference number: I52104; MUID:89024984; PMID:2460050  
 A:Accession: I52104  
 A>Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-217 <ICO>  
 A:Cross-references: GB:M38451; NID:q183179; L:V:AAA35891.1; PID:q183180  
 R:Frankenne, F.; Scippo, M.L.; Van Beeumen, J.; Igout, A.; Henneen, G.  
 J. Clin. Endocrinol. Metab. 71: 15-18, 1990  
 A:Title: Identification of placental human growth hormone as the growth hormone-V gene  
 A:Reference number: A60711; MUID:90317018; PMID:2196278  
 A:Accession: A60711  
 A:Molecule type: protein  
 A:Residues: 27-44; 46-57 <FRA>  
 A:Experimental source: tissue placenta  
 A:Note: partial glycosylation was demonstrated by lectin binding  
 C:Comment: This gene is expressed by the placenta.  
 C:Genetics:  
 A:Gene: GH2; GH2  
 A:Cross-references: GDB:119943; OMIM:139240  
 A:Map position: 17q22-17q24  
 A:Introns: 4/1; 57/3; 97/3; 152/3  
 C:Superfamily: prolactin  
 C:Keywords: alternative splicing; glycoprotein; hormone; placenta  
 F:1-26/domain: signal sequence #status predicted <SIG>  
 F:27-217/product: somatotropin 2, long splice form #status predicted <SIG>  
 F:27-57.73-217/product: somatotropin 2, short splice form #status predicted <SIG>  
 F:79-191,208-215/disulfide bonds: #status predicted  
 F:166/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 Query Match 89.8%; Score 422; DB 1; Length 217;  
 Best Local Similarity 92.3%; Pred. No. 5.9e-38;  
 Matches 84; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 FPTPLSLFDNAMLRAHRLHQLAFDYQEFEEAYIPKEOKYSLFONPOTSLSFSSEIPT 61  
 DB 27 FPTPLSLFDNAMLRAHRLHQLAFDYQEFEEAYIPKEOKYSLFONPOTSLSFSSEIPT 86  
 QY 62 PSNREETOOKSNLELLRISLLLIQSLEPVO 92  
 DB 87 PSNRVKTOOKSNLELLRISLLLIQSLEPVO 117

## RESULT 4

## STHUV2

somatotropin 2 precursor, splice form 2 - human  
 N:Alternate names: growth hormone variant-2; placental somatotropin form 2  
 C:Species: Homo sapiens (man)  
 C>Date: 30-Sep-1989 #sequence\_revision 10-Feb-1995 #text\_change 02-Sep-1997  
 C:Accession: A28072  
 R:Cooke, N.E.; Kay, J.; Emery, J.G.; Liebhauer, S.A.  
 J. Biol. Chem. 263: 9001-9006, 1988  
 A:Title: Two distinct species of human growth hormone variant mRNA in the human placenta  
 A:Reference number: A92725; MUID:88243759; PMID:3379057  
 A:Accession: A28072  
 A:Molecule type: mRNA  
 A:Residues: 1-256 <COO>  
 A:Note: an alternative splice junction for intron 4 is used  
 C:Genetics:  
 A:Gene: GH2; GH2  
 A:Cross-references: GDB:119943; OMIM:139240  
 A:Map position: 17q22-17q24  
 A:Introns: 4/1; 57/3; 97/3; 152/3  
 C:Superfamily: prolactin  
 C:Keywords: alternative splicing; hormone; placenta  
 F:1-256/domain: signal sequence #status predicted <SIG>  
 F:27-256/product: somatotropin 2 splice form 2 #status predicted <MAI>  
 Query Match 89.8%; Score 422; DB 1; Length 256;  
 Best Local Similarity 92.3%; Pred. No. 6,7e-38;  
 Matches 84; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 FPTPLSLFDNAMLRAHRLHQLAFDYQEFEEAYIPKEOKYSLFONPOTSLSFSSEIPT 61  
 DB 27 FPTPLSLFDNAMLRAHRLHQLAFDYQEFEEAYIPKEOKYSLFONPOTSLSFSSEIPT 86  
 QY 62 PSNREETOOKSNLELLRISLLLIQSLEPVO 92  
 DB 87 PSNRVKTOOKSNLELLRISLLLIQSLEPVO 117

## RESULT 5

## 167411

somatotropin - rhesus macaque  
 N:Alternate names: growth hormone  
 C:Species: Macaca mulatta (rhesus macaque)  
 C>Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 16-Jul-1999  
 C:Accession: I67411  
 R:Colos, J.G.; Durning, M.; Fisher, J.M.; Fowler, P.D.  
 Endocrinology 133: 1744-1752, 1993  
 A:Title: Cloning of four growth hormone/chorionic somatomammotropin related comple  
 A:Reference number: I53267; MUID:94008724; PMID:8404617  
 A:Accession: I67411  
 A>Status: preliminary; translated from GB/EMBL/GenBank  
 A:Molecule type: mRNA  
 A:Residues: 1-217 <RES>  
 A:Cross-references: GB:I16555; NID:q293116; PIDN:AAA20180.1; PID:q293117  
 C:Superfamily: prolactin  
 Query Match 85.5%; Score 402; DB 2; Length 217;  
 Best Local Similarity 85.7%; Pred. No. 7.8e-46;  
 Matches 78; Conservative 6; Mismatches 7; Indels 0; Gaps 0;  
 QY 2 FPTPLSLFDNAMLRAHRLHQLAFDYQEFEEAYIPKEOKYSLFONPOTSLSFSSEIPT 61  
 DB 27 FPTPLSLFDNAMLRAHRLHQLAFDYQEFEEAYIPKEOKYSLFONPOTSLSFSSEIPT 86  
 QY 62 PSNREETOOKSNLELLRISLLLIQSLEPVO 92  
 DB 87 PSNREETOOKSNLELLRISLLLIQSLEPVO 117

## RESULT 6

## 167409

chorionic somatomammotropin-3 - rhesus macaque

C:Species: Macaca mulatta (rhesus macaque)  
 C:Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 16-Jul-1999  
 C:Accession: 167409  
 R:Colos, T.G.; Durning, M.; Fisher, J.M.; Fowler, P.D.  
 Endocrinology 133, 1744-1752, 1993  
 A:Title: Cloning of four growth hormone/chorionic somatomotropin-related complementary DNAs from rhesus macaque  
 A:Reference number: 153267; MUID:94008724; PMID:8404617  
 A:Accession: 167409  
 A:Status: preliminary; translated from GR/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-217 <RES>  
 A:Cross-references: GB:L16554; NID:q293112; PIDN:AAA18841.1; PID:q293113  
 C:Superfamily: prolactin

Query Match 84.1%; Score 396; DB 2; Length 217;  
 Best Local Similarity 82.2%; Pred. No. 3.4e-35;  
 Matches 74; Conservative 11; Mismatches 5; Indels 0; Gaps 0;  
 QY 3 PTPLSLRFDNMLRAHRLHQLAFDTYQFEFAYIPKQKYSFLONPOTSLSFSESIPTP 62  
 DB 28 PSVPLSKLFDMIMQAHRLHQLAFDTYQFEFAYIPKQKYSFLONPOTSLSFSESIPTP 87  
 QY 63 SNRETOQKSNLELLRLISLLIQSWLEPVQ 92  
 DB 88 SNRETOQKSNLELLRLISLLIQSWLEPVQ 112

RESULT 7  
 167408  
 chorionic somatomotropin-2 - rhesus macaque (fragment)  
 C:Species: Macaca mulatta (rhesus macaque)  
 C:Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 16-Jul-1999  
 C:Accession: 167408  
 R:Colos, T.G.; Durning, M.; Fisher, J.M.; Fowler, P.D.  
 Endocrinology 133, 1744-1752, 1993  
 A:Title: Cloning of four growth hormone/chorionic somatomotropin-related complementary DNAs from rhesus macaque  
 A:Reference number: 153267; MUID:94008724; PMID:8404617  
 A:Accession: 167408  
 A:Status: preliminary; translated from GR/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-212 <RES>  
 A:Cross-references: GB:L16553; NID:q293110; PIDN:AAA18840.1; PID:q293111;  
 C:Superfamily: prolactin

Query Match 84.1%; Score 396; DB 2; Length 212;  
 Best Local Similarity 82.2%; Pred. No. 3.4e-35;  
 Matches 74; Conservative 11; Mismatches 5; Indels 0; Gaps 0;  
 QY 3 PTPLSLRFDNMLRAHRLHQLAFDTYQFEFAYIPKQKYSFLONPOTSLSFSESIPTP 62  
 DB 23 PSVPLSKLFDMIMQAHRLHQLAFDTYQFEFAYIPKQKYSFLONPOTSLSFSESIPTP 87  
 QY 63 SNRETOQKSNLELLRLISLLIQSWLEPVQ 92  
 DB 88 SNRETOQKSNLELLRLISLLIQSWLEPVQ 112

RESULT 6  
 153267  
 chorionic somatomotropin-1 - rhesus macaque  
 C:Species: Macaca mulatta (rhesus macaque)  
 C:Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 16-Jul-1999  
 C:Accession: 153267  
 R:Colos, T.G.; Durning, M.; Fisher, J.M.; Fowler, P.D.  
 Endocrinology 133, 1744-1752, 1993  
 A:Title: Cloning of four growth hormone/chorionic somatomotropin-related complementary DNAs from rhesus macaque  
 A:Reference number: 153267; MUID:94008724; PMID:8404617  
 A:Accession: 153267  
 A:Status: preliminary; translated from GR/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-217 <RES>  
 A:Cross-references: GB:L16552; NID:q293108; PIDN:AAA18839.1; PID:q293109  
 C:Superfamily: prolactin

Query Match 84.1%; Score 396; DB 2; Length 217;  
 Best Local Similarity 82.2%; Pred. No. 3.4e-35;  
 Matches 74; Conservative 11; Mismatches 5; Indels 0; Gaps 0;  
 QY 3 PTPLSLRFDNMLRAHRLHQLAFDTYQFEFAYIPKQKYSFLONPOTSLSFSESIPTP 62  
 DB 28 PSVPLSKLFDMIMQAHRLHQLAFDTYQFEFAYIPKQKYSFLONPOTSLSFSESIPTP 87  
 QY 63 SNRETOQKSNLELLRLISLLIQSWLEPVQ 92  
 DB 88 SNRETOQKSNLELLRLISLLIQSWLEPVQ 117

RESULT 9  
 153192  
 chorionic somatomotropin A precursor (validated) - human  
 C:Species: Homo sapiens (man)  
 C:Date: 23-Oct-1981 #sequence\_revision 23-Oct-1981 #text\_change 08-Dec-2000  
 C:Accession: C32435; A94422; 152342; A93833; A93192; A90054; A94427; A61283; 155229;  
 R:Chen, E.Y.-J.; Liang, Y.-C.; Smith, D.H.; Barrera-Saldana, R.A.; Gellinas, R.E.; Seeburg,  
 Genomics 4, 479-497, 1989  
 A:Title: The human growth hormone locus: nucleotide sequence, biology, and evolution  
 A:Reference number: A32435; MUID:89407277; PMID:2744760  
 A:Accession: C32435  
 A:Molecule type: DNA  
 A:Residues: 1-217 <CHE>  
 A:Cross-references: GB:004271; NID:q183148; PIDN:AAA52551.1; PID:q183151  
 R:Goodman, H.M.; DeNoto, F.; Fiddes, J.C.; Hillebrand, R.A.; Papo, G.S.; Smith, S.; T  
 in Mobilization and Reassembly of Genetic Information, Scott, W.A.; Werner, R.; Josep  
 A:Reference number: A94422  
 A:Accession: A94422  
 A:Molecule type: mRNA  
 A:Residues: 1-217 <GCO>  
 R:Tanaka, M.; Masuda, N.; Watahiki, M.; Yamakawa, M.; Shimizu, K.; Nagai, J.; Nakash  
 Biochem. Int. 16, 287-292, 1988  
 A:Title: cDNA cloning of human chorionic somatomotropin-1 mRNA whose transcription  
 A:Reference number: 152342; MUID:88209096; PMID:2815050  
 A:Accession: 152342  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-3 <TAN>  
 A:Cross-references: GB:M35419; NID:q506422  
 R:Sherwood, L.M.; Burstein, Y.; Schechter, I.  
 Proc. Natl. Acad. Sci. U.S.A. 76, 3819-3823, 1979  
 A:Title: Primary structure of the NH-2-terminal extra piece of the precursor to human  
 A:Reference number: A93833; MUID:80034970; PMID:291043  
 A:Accession: A93833  
 A:Molecule type: protein  
 A:Residues: 1-3-26 <SHE>  
 A:Experimental source: placenta  
 R:Shine, J.; Seeburg, P.H.; Martial, J.A.; Baxter, J.D.; Goodman, H.M.  
 Nature 270, 494-499, 1977  
 A:Title: Construction and analysis of recombinant DNA for human chorionic somatomamm  
 A:Reference number: A93192; MUID:78071761; PMID:593368  
 A:Accession: A93192  
 A:Molecule type: DNA  
 A:Residues: 50-217 <SHI>  
 A:Experimental source: placenta  
 R:Li, C.H.; Dixon, J.S.; Chung, D.  
 Arch. Biochem. Biophys. 155, 95-110, 1973  
 A:Title: Amino acid sequence of human chorionic somatomammotropin.  
 A:Reference number: A90054; MUID:73201971; PMID:4712450  
 A:Accession: A90054  
 A:Molecule type: protein  
 A:Residues: 27-217 <LIC>  
 A:Experimental source: placenta  
 R:Niall, H.D.  
 in Prolactin and Carcinogenesis, Proc. Fourth Tenovus Workshop Prolactin, Griffiths,  
 A:Title: The chemistry of the human lactogenic hormones.  
 A:Reference number: A94427  
 A:Accession: A94427

```

A: Molecule type: protein
A: Residues: 27-217 <NIA>
A: Experimental source: placenta
R: Nic A Baird, N.; Tipton, K.F.
Biochem. Soc. Trans. 19, 205, 1991
A: Title: Catechol-O-methyltransferase from human placenta: purification and some properties
A: Reference number: A61283; MUID: 91244006; PMID: 2037148
A: Accession: A61283
A: Molecule type: protein
A: Residues: 27-46 <NIC>
A: Note: chorionamniotropin apparently copurified with placental catechol-O-methyltransferase
R: Sherwood, L.M.; Handwerker, S.; McLaurin, W.D.; Lanner, M.
Nature New Biol. 233, 59-61, 1971
A: Title: Amino-acid sequence of human placental lactogen
A: Reference number: A93401; MUID: 72016313; PMID: 5286363
A: Contents: annotation
R: Sherwood, L.M.; Handwerker, S.; McLaurin, W.D.; Lanner, M.
Nature New Biol. 233, 59-61, 1971
A: Reference number: A93401
A: Contents: annotation
R: Schneider, A.B.; Kowalski, K.; Russell, J.; Sherwood, L.M.
J. Biol. Chem. 254, 3782-3787, 1979
A: Title: Identification of the interchain disulfide bonds of dimeric human placental lactogen
A: Reference number: A92251; MUID: 79173081; PMID: 438159
A: Contents: annotation; dimeric disulfide bonds
R: Selby, M.J.; Barta, A.; Baxter, J.D.; Bell, G.I.; Eberhardt, N.L.
J. Biol. Chem. 259, 13131-13138, 1984
A: Title: Analysis of a major human chorionic somatomammotropin gene. Evidence for two functional alleles
A: Reference number: 155229; MUID: 85030426; PMID: 6208192
A: Accession: 155229
A: Status: translated from GB/EMBL/DBJ
A: Molecule type: DNA
A: Residues: 1-217 <RES>
A: Cross-references: GB:K02401; MUID: q181120; PID: AAA52115.1; PID: q181121
R: Seeburg, P.H.; Shine, J.; Martial, J.A.; Ullrich, A.; Goodman, R.
Trans. Assoc. Am. Physicians 90, 109-116, 1977
A: Title: Nucleotide sequence of a human gene coding for a polypeptide hormone.
A: Reference number: 159658; MUID: 78160787; PMID: 611657
A: Accession: 159658
A: Status: translated from GB/EMBL/DBJ
A: Molecule type: mRNA
A: Residues: 169-217 <RES2>
A: Cross-references: GB:M25118; MUID: q181124; PID: AAA5721.1; PID: q181125
A: Gene: GDB:CSH1
A: Cross-references: GDB:119084; UMIN:152200
A: Map position: 17q22-17q24
A: Introns: 4/1; 57/3; 97/3; 152/4
A: Superfamily: prolactin
C: Keywords: hormone; placenta
F: 1-26/domain: signal sequence #status experimental <SIG>
F: 27-217/product: chorionamniotropin A #status experimental <MAT>
F: 79-191/disulfide bonds: #status experimental
F: 208-215/disulfide bonds: (in monomeric form) #status experimental
F: 208/disulfide bonds: interchain (to 215 in dimeric form) #status experimental
F: 215/disulfide bonds: interchain (to 238 in dimeric form) #status experimental

Query Match 81.1% Score 381; DB 2; Length 217;
Best Local Similarity 82.0% Pred. No. 1.4e-33;
Matches 73; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

Qy 4 TPLSLRFDNMLRAHRLHQLADTYQEFEEYIPKQKYSFLQNQTSLSFSISPTPS 63
Db 29 TVPLSLRFDHMLQAHRHQLADTYQEFEEYIPKQKYSFLQDSQTSFCFSDSISPTPS 88

Qy 64 NREETOQKSNLELLRISLLLSLQSLWLEPVQ 92
Db 89 NMEETOQKSNLELLRISLLLSLQSLWLEPVQ 117

RESULT 10
E32435
chorionamniotropin B precursor - human

A: Alternate names: chorionic somatomammotropin 2
C: Species: Homo sapiens (man)
C: Date: 29-Dec-1989 #sequence_revision 29-Dec-1989 #text_change 16-Jul-1999
C: Accession: E32435
R: Chen, E.Y.; Liao, Y.C.; Smith, D.H.; Barrera-Saldana, H.A.; Gelinas, R.E.; Seeburg, P.H.
Genomics 4, 475-487, 1989
A: Title: The human growth hormone locus: nucleotide sequence, biology, and evolution
A: Reference number: A32435; MUID: 85307277; PMID: 2744760
A: Accession: E32435
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-217 <CHG>
A: Cross-references: GB:JC3071; MUID: g183148; PID: AAA52553.1; PID: g183153
C: Genetics:
A: Gene: GDB:CSH2
A: Cross-references: GDB:119613; UMIN:118820
A: Map position: 17q22-17q24
C: Superfamily: prolactin

Query Match 81.1% Score 381; DB 2; Length 217;
Best Local Similarity 82.0% Pred. No. 1.4e-33;
Matches 73; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

Qy 4 TPLSLRFDNMLRAHRLHQLADTYQEFEEYIPKQKYSFLQNQTSLSFSISPTPS 63
Db 29 TVPLSLRFDHMLQAHRHQLADTYQEFEEYIPKQKYSFLQDSQTSFCFSDSISPTPS 88

Qy 64 NREETOQKSNLELLRISLLLSLQSLWLEPVQ 92
Db 89 NMEETOQKSNLELLRISLLLSLQSLWLEPVQ 117

RESULT 11
A26449
chorionamniotropin precursor (allele hCS-3) - human
C: Species: Homo sapiens (man)
C: Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 28-Jul-1995
C: Accession: A26449
R: Hirt, H.; Kimmelman, J.; Birnbaum, M.J.; Chen, E.Y.; Seeburg, P.H.; Eberhardt, N.L.
DNA 6, 59-70, 1987
A: Title: The human growth hormone gene locus: structure, evolution, and allelic variation
A: Reference number: A26449; MUID: 87161235; PMID: 3030680
A: Accession: A26449
A: Molecule type: DNA
A: Residues: 1-215 <HIR>
C: Superfamily: prolactin
F: 1-26/domain: signal sequence #status predicted <SIG>
F: 27-215/product: chorionamniotropin, hCS-3 allele #status predicted <MAT>

Query Match 76.5% Score 359.5; DB 2; Length 215;
Best Local Similarity 80.5% Pred. No. 2.9e-31;
Matches 70; Conservative 8; Mismatches 8; Indels 1; Gaps 1;

Qy 4 TPLSLRFDNMLRAHRLHQLADTYQEFEEYIPKQKYSFLQNQTSLSFSISPTPS 63
Db 29 TVPLSLRFDHMLQAHRHQLADTYQEFEEYIPKQKYSFLQDSQTSFCFSDSISPTPS 88

Qy 64 NREETOQKSNLELLRISLLLSLQSLWLEPVQ 90
Db 89 NMEETOQKSNLELLRISLLLSLQSLWLEPVQ 114

RESULT 12
B49159
somatotropin - golden hamster
A: Alternate names: growth hormone
C: Species: Mesocricetus auratus (golden hamster)
C: Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C: Accession: B49159
R: Southard, J.N.; Sanchez-Jimenez, F.; Campbell, G.T.; Talamantes, F.
Endocrinology 129, 2965-2971, 1991
A: Title: Sequence and expression of hamster prolactin and growth hormone messenger RNA
A: Reference number: A49159; MUID: 92063850; PMID: 1954881

```

A:Accession: B49159  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-216 <SD>  
 A:Cross-references: GB:S66299; NID:q233355; PIDN:AA20368.1; PID:q233346  
 A:Note: sequence extracted from NCBI backbone (NCBIN:66299, NCBIPI:66300)  
 C:Superfamily: prolactin

Query Match 66.1%; Score 310.5; DB 2; Length 216;  
 Best local similarity 67.0%; Pred. No. 5.5e-26;  
 Matches 61; Conservative 13; Mismatches 16; Indels 1; Gaps 1;

QY 2 FTTPLSLFNFAMLRHRLHQLAFDTYORFEAYIPKEQKYSFLQNPQTSLSFSSEIPT 61  
 11 11111 11111 11111 11111 11111 11111 11111 11111 11111  
 Db 27 FPMPLSLFANAVLRAQHLHQLAADTYKEFERAYIPGQRYIS-IONAQTAFSEIPTA 85  
 11 11111 11111 11111 11111 11111 11111 11111 11111 11111

QY 62 PSNRRETOOKSNLELRISLLLIOSWLEPVQ 92  
 11 11111 11111 11111 11111 11111 11111 11111 11111 11111  
 Db 86 PTGKEAQRSDMELLRFSLILLIOSWLGVPQ 116  
 11 11111 11111 11111 11111 11111 11111 11111 11111 11111

RESULT 13  
 STMS  
 somatotropin - sei whale  
 N:Alternate names: growth hormone  
 C:Species: Balanoptera borealis (sei whale)  
 C>Date: 07-May-1993 #sequence\_revision 07-May-1993 #text\_change 07-May-1993  
 C:Accession: P00140  
 R:Yudkev, N.A.; Pankov, Y.A.; Bulatov, A.A.; Osipova, I.A.  
 Biochimica 47, 1059-1069, 1982  
 A:Title: Amino acid sequence of sei whale somatotropin.  
 A:Reference number: P00140; MUID:84000569; PMID:7115814  
 A:Accession: P00140  
 A:Molecule type: protein  
 A:Residues: 1-190 <SD>  
 A:Note: article in Russian with English abstract  
 C:Superfamily: prolactin  
 C:Keywords: growth factor; hormone  
 F:52-163,180-188/Disulfide bonds: #status predicted

Query Match 65.4%; Score 307.5; DB 2; Length 190;  
 Best local similarity 67.0%; Pred. No. 9.8e-26;  
 Matches 61; Conservative 14; Mismatches 15; Indels 1; Gaps 1;

QY 2 FTTPLSLFNFAMLRHRLHQLAFDTYORFEAYIPKEQKYSFLQNPQTSLSFSSEIPT 61  
 11 11111 11111 11111 11111 11111 11111 11111 11111 11111  
 Db 1 FPMPLSLFANAVLRAQHLHQLAADTYKEFERAYIPGQRY-FLNAQSTQFSEIPT 59  
 11 11111 11111 11111 11111 11111 11111 11111 11111 11111

QY 62 PSNRRETOOKSNLELRISLLLIOSWLEPVQ 92  
 11 11111 11111 11111 11111 11111 11111 11111 11111 11111  
 Db 60 PANKDEAQRSDMELLRFSLILLIOSWLGVPQ 90  
 11 11111 11111 11111 11111 11111 11111 11111 11111 11111

RESULT 14  
 STMS  
 somatotropin precursor - mouse  
 N:Alternate names: growth hormone  
 C:Species: Mus musculus (house mouse)  
 C>Date: 10-Sep-1987 #sequence\_revision 10-Sep-1987 #text\_change 28-May-1999  
 C:Accession: B21911  
 R:Linzer, D.L.H.; Talamantes, F.  
 J. Biol. Chem. 260, 9574-9579, 1985  
 A:Title: Nucleotide sequence of mouse prolactin and growth hormone mRNAs and expression  
 A:Reference number: A92548; MUID:85261358; PMID:2491252  
 A:Accession: B21911  
 A:Molecule type: mRNA  
 A:Residues: 1-216 <LIN>  
 A:Cross-references: GB:X02891; NID:q51067; PIDN:CAA26650.1; PID:q51068  
 C:Superfamily: prolactin  
 C:Keywords: anterior pituitary; growth factor; hormone  
 F:1-216/Domain: signal sequence #status predicted <SIG>  
 F:27-216/Product: somatotropin #status predicted <STN>  
 F:78-189,206-214/Disulfide bonds: #status predicted

Query Match 64.8%; Score 304.5; DB 1; Length 216;  
 Best local similarity 64.8%; Pred. No. 2.4e-25;  
 Matches 59; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

QY 2 FTTPLSLFNFAMLRHRLHQLAFDTYORFEAYIPKEQKYSFLQNPQTSLSFSSEIPT 61  
 11 11111 11111 11111 11111 11111 11111 11111 11111 11111  
 Db 27 FPMPLSLFANAVLRAQHLHQLAADTYKEFERAYIPGQRYIS-IONAQAACFSEIPTA 85  
 11 11111 11111 11111 11111 11111 11111 11111 11111 11111

QY 62 PSNRRETOOKSNLELRISLLLIOSWLEPVQ 92  
 11 11111 11111 11111 11111 11111 11111 11111 11111 11111  
 Db 86 PTGKEAQRSDMELLRFSLILLIOSWLGVPQ 116  
 11 11111 11111 11111 11111 11111 11111 11111 11111 11111

RESULT 15  
 STMS  
 somatotropin - horse  
 N:Alternate names: growth hormone  
 C:Species: Equus caballus (domestic horse)  
 C>Date: 13-Jul-1981 #sequence\_revision 13-Jul-1981 #text\_change 23-Aug-1996  
 C:Accession: A91772; A91883; A91884; A92440; A92514  
 R:Zakin, M.M.; Poskus, E.; Bellach, J.M.; Paladini, A.C.; Santome, J.A.  
 Int. J. Pept. Protein Res. 6, 435-444, 1976  
 A:Title: Primary structure of equine growth hormone.  
 A:Reference number: A91772; MUID:77065419; PMID:966151  
 A:Accession: A91772  
 A:Molecule type: protein  
 A:Residues: 1-190 <ZAK>  
 R:Zakin, M.M.; Poskus, E.; Bellach, J.M.; Paladini, A.C.; Santome, J.A.  
 FEBS Lett. 16, 353-355, 1973  
 A:Title: The amino acid sequence of equine growth hormone.  
 A:Reference number: A91395; MUID:74020362; PMID:4747849  
 A:Accession: A91395  
 A:Molecule type: protein  
 A:Residues: 1-190 <ZAK>  
 R:Zakin, M.M.; Poskus, E.; Bellach, J.M.; Paladini, A.C.; Santome, J.A.  
 FEBS Lett. 25, 77-82, 1972  
 A:Title: Amino acid sequences around the cysteine residues in equine growth hormone.  
 A:Reference number: A91383  
 A:Accession: A91383  
 A:Molecule type: protein  
 A:Residues: 42-69;157-190 <ZAK>  
 R:Oliver, L.; Hartree, A.S.  
 Biochem. J. 109, 19-24, 1968  
 A:Title: Amino acid sequences around the cysteine residues in horse growth hormone.  
 A:Reference number: A90240; MUID:69368390; PMID:4876103  
 A:Accession: A90240  
 A:Molecule type: protein  
 A:Residues: 176-190 <OLI>  
 C:Superfamily: prolactin  
 C:Keywords: hormone; pituitary  
 F:52-163,180-188/Disulfide bonds: #status experimental

Query Match 64.4%; Score 302.5; DB 1; Length 190;  
 Best local similarity 64.8%; Pred. No. 3.4e-25;  
 Matches 59; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

QY 2 FTTPLSLFNFAMLRHRLHQLAFDTYORFEAYIPKEQKYSFLQNPQTSLSFSSEIPT 61  
 11 11111 11111 11111 11111 11111 11111 11111 11111 11111  
 Db 1 FPMPLSLFANAVLRAQHLHQLAADTYKEFERAYIPGQRYIS-IONAQAACFSEIPTA 59  
 11 11111 11111 11111 11111 11111 11111 11111 11111 11111

QY 62 PSNRRETOOKSNLELRISLLLIOSWLEPVQ 92  
 11 11111 11111 11111 11111 11111 11111 11111 11111 11111  
 Db 60 PTGKEAQRSDMELLRFSLILLIOSWLGVPQ 90  
 11 11111 11111 11111 11111 11111 11111 11111 11111 11111

Search completed: September 15, 2003, 12:04:19  
 Job time : 16.1685 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:54:00 : Search time 8.57348 seconds  
(without alignments)  
504.633 Million cell updates/sec

Title: US-09-423-100-2

Perfect score: 470

Sequence: 1 MFPTIPLSLFDNAMI RAHR.....NLELRISLLIQSWLPIVQ 52

Scoring table: BLOSUM62  
Gap: 10.0, Gap-ext 0.7

Searched: 127865 seqs, 47,463,065 total kbps

Total number of hits satisfying chosen parameters: 127865

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Swissprot\_41.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	46.0	97.9	217	1	SOMA_HUMAN
2	46.0	97.9	217	1	SOMA_MACHU
3	46.0	97.9	217	1	SOMA_PANTR
4	44.4	92.1	217	1	SOMA_SALIB
5	43.2	91.7	217	1	SOMA_CALJA
6	43.0	91.5	217	1	SOMA_PANTR
7	42.2	89.8	217	1	SOMA_HUMAN
8	39.6	84.3	217	1	SOMA_MACHU
9	38.1	81.1	217	1	PLI_HUMAN
10	31.0	66.1	216	1	SOMA_MESAJ
11	30.7	65.4	190	1	SOMA_BALRC
12	30.4	64.8	216	1	SOMA_MOUSE
13	30.2	64.4	216	1	SOMA_HORSE
14	30.2	64.4	216	1	SOMA_RABIT
15	30.2	64.4	216	1	SOMA_RAT
16	30.2	64.4	217	1	SOMA_GALSF
17	30.2	64.4	217	1	SOMA_NICPY
18	30.1	64.1	190	1	SOMA_LOXAF
19	30.1	64.1	216	1	SOMA_CANFA
20	30.1	64.1	216	1	SOMA_FELCA
21	30.1	64.1	216	1	SOMA_PIG
22	29.9	63.7	216	1	SOMA_MUSVI
23	29.7	63.3	190	1	SOMA_LAMPA
24	29.5	62.9	190	1	SOMA_YULVU
25	29.1	62.0	215	1	SOMA_MONDO
26	29.1	62.0	215	1	SOMA_TRIVU
27	28.9	61.6	217	1	SOMA_BOVIN
28	28.9	61.6	217	1	SOMA_CEREL
29	28.9	61.6	217	1	SOMA_SHEEP
30	28.2	60.1	217	1	SOMA_BUBBU
31	27.8	59.3	216	1	SOMA_MELGA
32	27.5	58.6	216	1	SOMA_CHICK
33	27.4	58.4	217	1	SOMA_STRCA

34 272.5 58.0 190 1 SOMA\_CRONG  
35 268.5 57.1 191 1 SOMA\_CHEMY  
36 261 55.5 216 1 SOMA\_ANAPL  
37 257.5 54.8 190 1 SOM1\_AC1GU  
38 257.5 54.8 190 1 SOM2\_AC1GU  
39 247.5 52.7 211 1 SOMA\_LEPOS  
40 239.5 51.0 214 1 SOMA\_XENLA  
41 238.5 50.7 215 1 SOMA\_RANCA  
42 226.5 48.2 213 1 SOMA\_BUFMA  
43 225.5 48.0 183 1 SOMA\_PRIGL  
44 219.5 46.7 208 1 SOMB\_XENLA  
45 218.5 46.5 206 1 SOMA\_PROAN

#### ALIGNMENTS

RESULT 1  
SOMA\_HUMAN SEABARDI PRS 217 AA  
AC P01241 Q24405 Q16641 Q98671 Q50M37 Q60M55  
DT 21-JUL-1986 (Ref: 01, Created)  
DT 01-MAR-1992 (Ref: 21, last sequence update)  
DE 15-SEP-2003 (Ref: 47, last annotation update)  
DE Somatotropin precursor r (Growth hormone) (GH) (GH-N) (Pituitary growth hormone) (Growth hormone 1).  
GN GH1  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE:80034477; PubMed:386281;  
RA Roskam W., Roudeon F.;  
RT "Molecular cloning and nucleotide sequence of the human growth hormone structural gene";  
RL Nucleic Acids Res. 7:305-320(1979).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE:79201293; PubMed:377496;  
RA Martial J.A., Haillet R.A., Baxter J.D., Goodman H.M.;  
RT "Human growth hormone: complementary DNA cloning and expression in bacteria";  
RL Science 205:602-607(1979).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND POSSIBLE ALTERNATIVE SPLICING.  
RX MEDLINE:82014939; PubMed:6269091;  
RA Denoto F.M., Moore D.D., Goodman H.M.;  
RT "Human growth hormone DNA sequence and mRNA structure: possible alternative splicing";  
RL Nucleic Acids Res. 9:3719-3730(1981).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE:83182010; PubMed:7169009;  
RA Seeburg P.H.;  
RT "The human growth hormone gene family: nucleotide sequences show recent divergence and predict a new polypeptide hormone";  
RL DNA 1:239-249(1982).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE:89307277; PubMed:2744760;  
RA Chen E.Y., Liao Y.C., Smith D.H., Barrera-Saldana H.A.,  
RA Gellinas R.E., Seeburg P.H.;  
RT "The human growth hormone locus: nucleotide sequence, biology, and evolution";  
RL Genomics 4:479-497(1989).  
RN [6]  
RP SEQUENCE FROM N.A. (ISOFORM 3).  
RT TISSUE-Pituitary;  
RA Gu J., Huang Q.-H., Li N., Xu S.-H., Han Z. G., Fu G., Chen Z.;  
RT "A novel gene expressed in human pituitary";  
RN Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases





```

RESULT 2
SOMA_MACMU STANDARD: PWT: 217 A.
ID ID SOMA MACMU
AC E33034,
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1994 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Somatotropin precursor (growth hormone) (GH-N) (Pituitary growth hormone) (Growth hormone 1).
DE DE GH1.
OS Macaca mulatta (Rhesus macaque).
OS Fukuyuta; Metazoa; Chordata; Granata; :obrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhin Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9544;
RN {}
RN SEQUENCE FROM N.A.
RP MEDLINE=94008724; PubMed=8404617;
RA Golos I.G., Durning M., Fisher J.M., Fowler P.D.;
RT "Cloning of four growth hormone/chorionic somatomotropin-related complementary deoxyribonucleic acids differentially expressed during pregnancy in the rhesus monkey placenta.";
RT Endocrinology 133:1744-1752(1993);
RL [2]
RN SEQUENCE OF 27-217.
RP MEDLINE=B6129460; PubMed=1080959;
RA Li C.H., Chung D., Lahm H.W., Stein S.;
RT "The primary structure of monkey pituitary growth hormone.";
RL Arch. Biochem. Biophys. 245:287-291(1986)
CC FUNCTION: Plays an important role in growth control. Its major role in stimulating body growth is to stimulate the liver and other tissues to secrete IGF-1. It stimulates both the differentiation and proliferation of myoblasts. It also stimulates amino acid uptake and protein synthesis in muscle and other tissues.
CC SUBCELLULAR LOCATION: Secreted.
CC SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC This SWISS-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/account/ or send an email to license@isb-sib.ch).
CC FMR1; L16556; AAA18842.1;
DR FIR; 167410; 167410.
DR MSP; P01241; IAXI.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
DR Hormone; Pituitary; Signal.
FT SIGNAL
FT 1
FT CHAIN 26
FT 27 217 SOMATOTROPIN.
FT DISULFID 79 191 BY SIMILARITY.
FT DISULFID 208 215 BY SIMILARITY.
FT CONFLICT 100 100 F > O (IN REF 2)

```

```

QY 62 PSNREETQOKSNLELLRISLLIQSWLEPVQ 2
DB 87 PSNREETQOKSNLELLRISLLIQSWLEPVQ 117

RESULT 4
SOMA_SA:IBB
ID SOMA_SA:IBB STANDARD; PRT: 217 AA.
AC P58343;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Somatotropin precursor (growth hormone).
GN
OS Saimiri boliviensis boliviensis (Bolivian squirrel monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.
OX NCBI_TaxID:9442;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE:21265430; PubMed:11371562;
RA Liu J.C., Makova K.D., Adkins R.M., Gibson S., Li W.H.;
RT "Epistatic evolution of growth hormone in primates and emergence of the
RT species specificity of human growth hormone receptor.*";
RL Mol. Biol. Evol. 18:945-953(2001)
CC
CC -!- FUNCTION: Plays an important role in growth control. Its major
CC role in stimulating body growth is to stimulate the liver and
CC other tissues to secrete IGF-1. It stimulates both the
CC differentiation and proliferation of myoblasts. It also stimulates
CC amino acid uptake and protein synthesis in muscle and other
CC tissues (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF349060; AAK62287.1;
CC InterPro: IPR001400; Somatotropin.
CC Pfam: PF00103; hormone; 1.
CC PRINTS: PF00836; SOMATOTROPIN.
CC PROSITE: PS00266; SOMATOTROPIN_1; 1.
CC PROSITE: PS00338; SOMATOTROPIN_2; 1.
CC Hormone; Pituitary; Signal.
CC SIGNAL: 1 26 BY SIMILARITY.
CC CHAIN: 27 217 SOMATOTROPIN.
CC DISULFID: 74 191 BY SIMILARITY.
CC DISULFID: 208 215 BY SIMILARITY.
CC SEQUENCE 217 AA: 24864 MW: 95152899.92997 CRR64;

Query Match 92.18; Score 433; DB 1; Length 217;
Best Local Similarity 91.28; Pred. No. 5,6e-38;
Matches 81; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 FPTPLSRLEFDNAMIKAHRLHQAFDTYQFFEAYIPKEOKYSFLNPQTSLSFSSEIPT 61
DB 27 FPTPLSRLEFDNAMIKAHRLHQAFDTYQFFEAYIPKEOKYSFLNPQTSLSFSSEIPT 86

QY 62 PSNREETQOKSNLELLRISLLIQSWLEPVQ
DB 87 PSNREETQOKSNLELLRISLLIQSWLEPVQ 7

RESULT 5
SOMA_CALJA
ID SOMA_CALJA STANDARD; PRT: 217 AA.
AC Q9GMB3;

```

```

DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Somatotropin precursor (Growth hormone).
GN
OS Callithrix jacchus (Common marmoset).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
OX Callithrix.
OX NCBI_TaxID:9483;
RN
RP SEQUENCE FROM N.A.
RX Wallis G.C., Wallis M.;
RT "Cloning and characterisation of a putative growth hormone encoding
RT gene from the marmoset (Callithrix jacchus).";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC
CC -!- FUNCTION: Plays an important role in growth control. Its major
CC role in stimulating body growth is to stimulate the liver and
CC other tissues to secrete IGF-1. It stimulates both the
CC differentiation and proliferation of myoblasts. It also stimulates
CC amino acid uptake and protein synthesis in muscle and other
CC tissues (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AJ297563; CAC03431.1;
CC HSP: P01241; I422.
CC InterPro: IPR001400; Somatotropin.
CC Pfam: PF00103; hormone; 1.
CC PROSITE: PS00266; SOMATOTROPIN_1; 1.
CC PROSITE: PS00338; SOMATOTROPIN_2; 1.
CC Hormone; Pituitary; Signal.
CC SIGNAL: 1 26 BY SIMILARITY.
CC CHAIN: 27 217 SOMATOTROPIN.
CC DISULFID: 79 191 BY SIMILARITY.
CC DISULFID: 208 215 BY SIMILARITY.
CC SEQUENCE 217 AA: 24959 MW: 2102151A12CE6192 CRR64;

Query Match 91.98; Score 432; DB 1; Length 217;
Best Local Similarity 91.28; Pred. No. 7.2e-48;
Matches 83; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 FPTPLSRLEFDNAMIKAHRLHQAFDTYQFFEAYIPKEOKYSFLNPQTSLSFSSEIPT 61
DB 27 FPTPLSRLEFDNAMIKAHRLHQAFDTYQFFEAYIPKEOKYSFLNPQTSLSFSSEIPT 86

QY 62 PSNREETQOKSNLELLRISLLIQSWLEPVQ 92
DB 87 PSNREETQOKSNLELLRISLLIQSWLEPVQ 117

RESULT 6
SOM2_PANTR
ID SOM2_PANTR STANDARD; PRT: 217 AA.
AC P58757;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Growth hormone variant precursor (GH-V) (Placenta-specific growth
DE hormone) (Growth hormone 2).
GN GH2.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
OX NCBI_TaxID:9598;

```





RT \*Analysis of a major human chorionic somatomammotropin gene. Evidence  
 RT for two functional promoter elements.";  
 RL J. Biol. Chem. 259:13131-13138(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A. (GENE CSN2)  
 RX MEDLINE=87161235; PubMed=3030680;  
 RA Hirt H., Kimmelmann J., Birnbaum M.J., Chen E.Y., Seeburg P.H.,  
 RA Eberhardt N.L., Barta A.;  
 RT "The human growth hormone gene locus: structure, evolution, and  
 RT allelic variations.";  
 RL DNA 6:59-70(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=83160916; PubMed=63000056;  
 RA Barrera-Saladana H.A., Seeburg P.H., Saunders G.F.;  
 RT "Two structurally different genes produce the same secreted human  
 RT placental lactogen hormone.";  
 RL J. Biol. Chem. 258:3787-3794(1983).  
 RN [4]  
 RP SEQUENCE FROM N.A. (GENE CSN1 AND CSN2)  
 RX MEDLINE=89407277; PubMed=2744760;  
 RA Chen E.Y., Liao Y.C., Smith D.H., Barrera-Saladana H.A., Golligorsky E.,  
 RA Seeburg P.H.;  
 RT "The human growth hormone locus: nucleotide sequence, biology, and  
 RT evolution.";  
 RL Genomics 4:479-497(1989).  
 RN [5]  
 RP SEQUENCE  
 RX MEDLINE=83482010; PubMed 7169009;  
 RA Seeburg P.H.;  
 RT "The human growth hormone gene family: nucleotide sequences show  
 RT recent divergence and predict a new polypeptide hormone.";  
 RL DNA 1:239-249(1982).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta, and Uterus;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen G.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
 RA Diatchenko L., Marusik K., Farmer A., Rubin G.M., Hong E.,  
 RA Skapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Soderstrom E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smolius D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [7]  
 RP SEQUENCE OF 50-217 FROM N.A.  
 RX MEDLINE=78071761; PubMed=593368;  
 RA Shino J., Seeburg P.H., Martial J.A., Baxter J.D., Goodman H.M.;  
 RT "Construction and analysis of recombinant DNA for human chorionic  
 RT somatomammotropin.";  
 RL Nature 270:494-499(1977).  
 RN [8]  
 RP SEQUENCE OF 27-217  
 RX MEDLINE=73201971; PubMed=4712450;  
 RA Li C.H., Dixon J.S., Chung D.;  
 RT "Amino acid sequence of human chorionic somatomammotropin.";  
 RL Arch. Biochem. Biophys. 155:95-110(1973).  
 RN [9]  
 RP SEQUENCE OF 27-117  
 RX MEDLINE=72016313; PubMed=5286363;

RA Sherwood L.M., Handwerger S., McLaurin W.D., Lanner M.;  
 RT "Amino-acid sequence of human placental lactogen.";  
 RL Nature New Biol. 233:59-61(1971).  
 RN [10]  
 RP ERRATUM.  
 RA Sherwood L.M., Handwerger S., McLaurin W.D., Lanner M.;  
 RL Nature New Biol. 235:64-64(1972).  
 RN [11]  
 RP INTERCHAIN DISULFIDE BONDS  
 RX MEDLINE=79173081; PubMed=438159;  
 RA Schneider A.B., Kowalski K., Russell J., Sherwood L.M.;  
 RT "Identification of the interchain disulfide bonds of dimeric human  
 RT placental lactogen.";  
 RL J. Biol. Chem. 254:3782-3787(1979).  
 CC - FUNCTION: SIMILAR TO THAT OF SOMATOTROPIN.  
 CC - SUBCELLULAR LOCATION: Secreted.  
 CC - MISCELLANEOUS: THE SEQUENCE OF CSN1 IS SHOWN.  
 CC - SIMILARITY: BELONGS TO THE SOMATOTROPIN/POLYGLUTAMIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: V00573; CAA23836.1;  
 DR EMBL: J00289; AAA98747.1;  
 DR EMBL: K02401; AAA52115.1;  
 DR EMBL: M15894; AAA52116.1;  
 DR EMBL: J03071; AAA52551.1;  
 DR EMBL: J00118; AAA98621.1;  
 DR EMBL: BC002717; AAH02717.1;  
 DR EMBL: BC005921; AAH05921.1;  
 DR EMBL: BC020756; AAH20756.1;  
 DR PIR: A26449; A26449.  
 DR PIR: C12435; LCHUC.  
 DR HSSP: P01241; 1A22.  
 DR Genew: HGNC:2440; CSN1.  
 DR Genew: HGNC:2441; CSN2.  
 DR MIM: 150200;  
 DR GO: GO:0007565; P:pregnancy; TAS.  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF0104; hormone\_1  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 KW Hormone; Placenta; Multigene family; Signal.  
 FT SIGNAL 1 25  
 FT CHAIN 27 217 LACTOGEN.  
 FT DISULFID 79 131  
 FT DISULFID 208 215  
 FT DISULFID 208 208  
 FT DISULFID 215 215  
 FT VARIANT 3 3  
 FT VARIANT 104 105  
 FT CONFLICT 84 84  
 FT CONFLICT 95 95  
 FT CONFLICT 116 116  
 FT CONFLICT 134 136  
 FT SEQUENCE 217 AA; 25020 MW; 235B0DC7A711F431 CRC64;  
 SO  
 Query Match 41.1%; Score 381; DB 1; Length 217;  
 Best Local Similarity 82.0%; Pred. No. 1.4e-32;  
 Matches 73; Conservative 8; Mismatches 8; Indels 0; Gaps 0;  
 QY 4 TIPLSRFDNMLRAHRLHQLAFDTYQFFEFAYIPKEQKYSFLQNPQTSIFSSEIPTPS 63  
 DB 29 TVPLSRFLDMLQAHRAHQLALDTYQEFRETYIPKQKYSFLHDSQTSFCFSDSIPTPS 88  
 QY 64 NRETOQSKNLELLRLISLLLIQSWLEPVQ 92



```

J. Biol. Chem. 260:9574-9579(1985).
[2]
SEQUENCE FROM N.A.
STRAN-ETIDU: TISSUE=Liver;
MEDLINE-96194803; PubMed-8647448;
RA Das P., Meyer L., Seyfert H.-M., Brockmann G., Schwerin M.:
RT "Structure of the growth hormone-encoding gene and its promoter in
RT mice.";
RL Gene 169:209-213(1996).
CC -1- FUNCTION: Plays an important role in growth control. Its major
CC role in stimulating body growth is to stimulate the liver and
CC other tissues to secrete IGF-1. It stimulates both the
CC differentiation and proliferation of myoblasts. It also stimulates
CC amino acid uptake and protein synthesis in muscle and other
CC tissues.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X02891; CAA26650.1;
DR EMBL: Z46663; CAA86658.1;
DR PIR: B23911; STMS.
DR HSSP: P01246; IBSI.
DR MGI: MGI:95707; Gh.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
DR Hormone: Pituitary; Signal.
KW SIGNAL 1 26 BY SIMILARITY.
FT CHAIN 27 216 SOMATOTROPIN.
FT DISULFID 78 189 BY SIMILARITY.
FT DISULFID 206 214 BY SIMILARITY.
SQ SEQUENCE 216 AA: 24716 MW: 986666 JAE25D65FC CRC64:
Query Match 64.8%; Score 302.5; DB 1; Length 216;
Best Local Similarity 64.8%; Pred. No. 1, 20-24;
Matches 59; Conservative 14; Mismatches 17; Indels 1; Gaps 1;
QY 2 FPTPLSRFDNAMLRAHLRHLQAFDTYQFFEFAYIPKEQKYSFLONPQTSLSFSSESIPT 61
Db 27 FPMPLSSLSFNAVLRAQHLRHLQAFDTYQFFEFAYIPKEQKYS-IGNAQAACFCSETIPA 85
QY 62 PSNREETQOKSNLELLRISLLLIQSWLEPVQ 6
Db 86 PTGKDEAQRSDMELLRFSLLIQSWLEPVQ 6
RESULT 13
SOMA_HORSE STANDARD; PRT: 216 AA.
AC P01245;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Somatotropin precursor (Growth hormone).
GN GH1.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RT TISSUE=Pituitary.
RX MEDLINE-94266171; PubMed-8206392;
RA Ascacio-Martinez J.A., Barrera-Saldana H.A.;

```

```

*Sequence of a cDNA encoding horse growth hormone.*;
Gene 143:299-300(1994).
[2]
SEQUENCE OF 27-216.
RX MEDLINE-77005410; PubMed-965151;
RA Zakin M.M., Poskus E., Langton A.A., Ferrara P., Santome J.A.;
RA Dellacha J.M., Paladini A.C.;
RT "Primary structure of equine growth hormone.";
RL Int. J. Pept. Protein Res. 8:435-444(1976).
CC [3]
CC PRELIMINARY SEQUENCE OF 27-216.
RX MEDLINE-7402362; PubMed-4747849;
RA Zakin M.M., Poskus E., Dellacha J.M., Paladini A.C., Santome J.A.;
RT "The amino acid sequence of equine growth hormone.";
RL FEBS Lett. 34:353-355(1973).
CC [4]
CC SEQUENCE OF 68 95 AND 183-216.
RA Zakin M.M., Poskus E., Dellacha J.M., Paladini A.C., Santome J.A.;
RT "Amino acid sequences around the cysteine residues in equine growth
RT hormone.";
RL FEBS Lett. 25:77-82(1972).
CC [5]
CC SEQUENCE OF 202-216.
RX MEDLINE-68368190; PubMed-4876100;
RA Oliver L., Hartree A.S.;
RT "Amino acid sequences around the cysteine residues in horse growth
RT hormone.";
RL Biochem. J. 109:19-24(1968).
CC -1- FUNCTION: Plays an important role in growth control. Its major
CC role in stimulating body growth is to stimulate the liver and
CC other tissues to secrete IGF-1. It stimulates both the
CC differentiation and proliferation of myoblasts. It also stimulates
CC amino acid uptake and protein synthesis in muscle and other
CC tissues.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U02929; AAA21027.1;
DR HSSP: P01246; IBSI.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
DR Hormone: Pituitary; Signal.
KW SIGNAL 1 26 SOMATOTROPIN.
FT CHAIN 27 216
FT DISULFID 78 189
FT DISULFID 206 214
SQ SEQUENCE 216 AA: 24423 MW: 17AB3173834E11AC CRC64:
Query Match 64.4%; Score 302.5; DB 1; Length 216;
Best Local Similarity 64.8%; Pred. No. 20-24;
Matches 59; Conservative 14; Mismatches 17; Indels 1; Gaps 1;
QY 2 FPTPLSRFDNAMLRAHLRHLQAFDTYQFFEFAYIPKEQKYSFLONPQTSLSFSSESIPT 61
Db 27 FPMPLSSLSFNAVLRAQHLRHLQAFDTYQFFEFAYIPKEQKYS-IGNAQAACFCSETIPA 85
QY 62 PSNREETQOKSNLELLRISLLLIQSWLEPVQ 92
Db 86 PTGKDEAQRSDMELLRFSLLIQSWLEPVQ 116
RESULT 14
SOMA_RABIT

```



ID	SOMA_RABIT	STANDARD:	PRT:	216 AA.
AC	P46407;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Somatotropin precursor (Growth hormone).			
GN	GHI			
OS	Oryctolagus cuniculus (Rabbit)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.			
OX	NB1_taxonomy:9986;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-New Zealand white;			
RA	MEDLINE:96011643; PubMed:7590276;			
RX	Wallis O C., Wallis M.;			
DE	*Cloning and Characterisation of the rabbit growth hormone encoding gene.;			
RT	Gene 163:253-256(1995).			
RL	-!- FUNCTION: Plays an important role in growth control. Its major role in stimulating body growth is to stimulate the liver and other tissues to secrete IGF-1. It stimulates both the differentiation and proliferation of myoblasts. It also stimulates amino acid uptake and protein synthesis in muscle and other tissues			
CC	-!- SUBCELLULAR LOCATION: Secreted.			
CC	-!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.			
CC	This SWISS PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> )			
CC	EMBL: Z38127; CAA86287.1;			
DR	PIR: S49483; S49483.			
DR	HSSP: P01246; IHST.			
DR	InterPro: IPR001400; Somatotropin.			
DR	Param: PF00103; hormone: 1.			
DR	PRINTS: PR00836; SOMATOTROPIN			
DR	PROSITE: PS00266; SOMATOTROPIN_1; 1			
DR	PROSITE: PS00338; SOMATOTROPIN_2; 1.			
KW	Hormone; Pituitary; Signal.			
FT	SIGNAL 1 26 POTENTIAL.			
FT	CHAIN 27 216 SOMATOTROPIN.			
FT	DISULFID 78 189 BY SIMILARITY.			
FT	DISULFID 206 214 BY SIMILARITY.			
FSO	SEQUENCE 216 AA; 2443 MW; 6EC1974R199FD75 CRC64;			
QY	Query Match 64.4%; Score 302.5; DR 1; Length 216;			
DB	Best Local Similarity 64.8%; Pred. No. 26-24;			
DB	Matches 59; Conservative 14; Mismatches 17; Indels 1; Gaps 1;			
QY	2 FTTPEKLFUNAMLRHRLHAFDTQGEFEAYIFKEQKYSFLQNPGTSLSFSESTPT 61 			
DB	27 FPAMPISLLFANAVLAQHHLKALVTKEFEAYIFEGQKYS LQAQAACEFTIPA 85			
QY	62 PSNKERTOOKSNLEELIKSLILLIQSKLEPVQ 92 			
DB	86 PTCKDAQQSDMELLRFSLILLIQSKMGVPQ 1-			
RESULT 15				
SOMA_RAT				
ID	SOMA_RAT	STANDARD:	PRT:	216 AA.
AC	P01244;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Somatotropin precursor (Growth hormone).			
GN	GHI OR GH.			





## RESULT 2

ID 014644 PRELIMINARY: PRT: 245 AA  
 AC 014644:  
 DT 01-JAN-1998 (TREMBlrel: 05, Created)  
 DT 01-JAN-1998 (TREMBlrel: 05, last sequence update)  
 DT 01-DEC-2001 (TREMBlrel: 19, last annotation update)  
 DE Placental growth hormone isoform hGH-V3 precursor  
 GH-V  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Term placenta;  
 RX MEDLINE=98374737; PubMed=9709463;  
 RA Bouzewska C., Carlsson B., Carlsson T., Clark R.  
 RT \*Cloning of two novel growth hormone transcripts expressed in human  
 placenta.\*  
 RL J. Clin. Endocrinol. Metab. 83:2878-2885(1998).  
 DR EMBL: AF006061; AAB71829.1;  
 DR HSSP: P01241; 1A22.  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone: 1.  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 KW Signal.  
 FT SIGNAL: 1 26 POTENTIAL.  
 SQ SEQUENCE 245 AA: 27101 MW: 140078607.49108 CRC64;  
 Query Match 89.8%; Score 422; ID 4; Length 245;  
 Best Local Similarity 92.3%; Pred. No. 1 245-39;  
 Matches 84; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 Qy 2 FPIPLSKLFDNMLRAHRLHQLAFDYQEFAYIPKQKYSFLONPQTSISFSESIPT 61  
 Db 27 FPIPLSKLFDNMLRAHRLHQLAFDYQEFAYIPKQKYSFLONPQTSISFSESIPT 61  
 Qy 52 PSNRETOQKSNLELLRLISLLLSQSWLEPVQ 92  
 Db 87 PSNRETOQKSNLELLRLISLLLSQSWLEPVQ 117

## RESULT 1

ID 007369 PRELIMINARY: PRT: 217 AA.  
 AC 007369:  
 DT 01-NOV-1996 (TREMBlrel: 01, Created)  
 DT 01-NOV-1996 (TREMBlrel: 01, last sequence update)  
 DT 01-MAR-2003 (TREMBlrel: 23, last annotation update)  
 DE Chorionic somatomotropin-3.  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;  
 OC Cercopitheciinae; Macaca.  
 OX NCBI\_TaxID=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Midpregnancy placenta;  
 RX MEDLINE=94008724; PubMed=8404617;  
 RA Golos T.G., Durnung M., Fisher J.M., Fowler P.D.;  
 RT \*Cloning of four growth hormone/chorionic somatomotropin-related  
 complementary deoxyribonucleic acids differentially expressed during  
 pregnancy in the rhesus monkey placenta.\*  
 RL Endocrinology 133:1744-1752(1993).  
 DR EMBL: L16554; AAA1884.1;  
 DR HSSP: P01241; 1AXI.  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone: 1.  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 SQ SEQUENCE 217 AA: 24874 MW: 11686AFDBA1B185 CRC64;

Query Match 84.5%; Score 397; DB 6; Length 217;  
 Best Local Similarity 83.3%; Pred. No. 6 9e-37;  
 Matches 75; Conservative 8; Mismatches 7; Indels 0; Gaps 0;  
 Qy 3 PTPLSKLFDNMLRAHRLHQLAFDYQEFAYIPKQKYSFLONPQTSISFSESIPT 62  
 Db 26 PSVPLSKLFDNMLRAHRLHQLAFDYQEFAYIPKQKYSFLONPQTSISFSESIPT 62  
 Qy 63 SNRETOQKSNLELLRLISLLLSQSWLEPVQ 92  
 Db 88 SNRETOQKSNLELLRLISLLLSQSWLEPVQ 117  
 RESULT 4  
 Q07368  
 ID 007368 PRELIMINARY: PRT: 212 AA.  
 AC 007368:  
 DT 01-NOV-1996 (TREMBlrel: 01, Created)  
 DT 01-NOV-1996 (TREMBlrel: 01, last sequence update)  
 DT 01-MAR-2003 (TREMBlrel: 23, last annotation update)  
 LE Somatotropin 2 precursor (Growth hormone 2) (Fragment).  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;  
 OC Cercopitheciinae; Macaca.  
 OX NCBI\_TaxID=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=94008724; PubMed=8404617;  
 RA Golos T.G., Durnung M., Fisher J.M., Fowler P.D.;  
 RT \*Cloning of four growth hormone/chorionic somatomotropin-related  
 complementary deoxyribonucleic acids differentially expressed during  
 pregnancy in the rhesus monkey placenta.\*  
 RL Endocrinology 133:1744-1752(1993).  
 DR EMBL: L16553; AAA1884.1;  
 DR HSSP: P01241; 1AXI.  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone: 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 212 AA: 24525 MW: 27BC911062566F5 CRC64;  
 Query Match 84.3%; Score 396; DB 6; Length 212;  
 Best Local Similarity 82.3%; Pred. No. 8 7e-47;  
 Matches 74; Conservative 11; Mismatches 5; Indels 0; Gaps 0;  
 Qy 3 PTPLSKLFDNMLRAHRLHQLAFDYQEFAYIPKQKYSFLONPQTSISFSESIPT 62  
 Db 23 PSVPLSKLFDNMLRAHRLHQLAFDYQEFAYIPKQKYSFLONPQTSISFSESIPT 62  
 Qy 63 SNRETOQKSNLELLRLISLLLSQSWLEPVQ 92  
 Db 83 SNLETOQKSNLELLRLISLLLSQSWLEPVQ 112  
 RESULT 5  
 Q07367  
 ID 007367 PRELIMINARY: PRT: 217 AA.  
 AC 007367:  
 DT 01-NOV-1996 (TREMBlrel: 01, Created)  
 DT 01-NOV-1996 (TREMBlrel: 01, last sequence update)  
 DT 01-MAR-2003 (TREMBlrel: 23, last annotation update)  
 DE Chorionic somatomotropin-1.  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;  
 OC Cercopitheciinae; Macaca.  
 OX NCBI\_TaxID=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Midpregnancy placenta;

```

DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
SO SEQUENCE 217 AA; 24994 MW; 39FAACDD86B2E951 CRC64;

Query Match 81.1%; Score 381; DB 4; Length 217;
Best Local Similarity 82.0%; Pred. No. 4.4e-35;
Matches 73; Conservative 8; Mismatches 8; Indels 0; Gaps 0.

QY 4 TIPSLSRFDNMLRAHRLHQAFYTYQEFKAYIPKQKYSFLQNPQTSLSFSSESIPTPS 63
DB 29 TVPLSRFLFDHMLQARHQAHLQADTYQEFERTYIPKQKYSFLHDSQTSFCSFSIPTPS 88
QY 64 NREETQKSNLELRISLLLIQSWLPEVQ 92
DB 89 NMETQKSNLELRISLLLIQSWEFPPVR 117

RESULT 7
QWRN25 PREINITIARY: PRT: 217 AA
AC QWRN25;
D: 01-MAR-2002 (TRMBLrel_20, Created)
DT 01-MAR-2002 (TRMBLrel_20, Last sequence update)
DT 03-OCT-2002 (TRMBLrel_22, Last annotation update)
DE Growth hormone.
DE GH-V.
GN Ateles geoffroyi (Black-handed spider monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Ateles.
OX NCBI_TaxID:9509;
RN [1]
RP SEQUENCE FROM N.A.
RA Revol A., Esquivel D., Santiago D., Barrera-Saldana H.;
RT "Independent duplication of the growth hormone gene in three
RT Anthropoid lineages.";
RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
RL EMBL: AF374235; AA:72287.1; -.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone.1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
SO SEQUENCE 217 AA; 25293 MW; 741745A1B75C053E CRC64;

Query Match 74.9%; Score 348; DB 6; Length 217;
Best Local Similarity 75.8%; Pred. No. 2.4e-31;
Matches 68; Conservative 8; Mismatches 14; Indels 0; Gaps 0.

QY 2 FPTPLSRFLFNLRAHRLHQAFYTYQEFKAYIPKQKYSFLQNPQTSLSFSSESIP 61
DB 27 FPRPLSRFLFDHMLRAHRLHQAFYTYQEFKAYIPKQKYSFLHDSQTSFCSFS 86
QY 62 PSNREETQKSNLELRISLLLIQSWLPEVQ 92
DB 87 PNKEEVLAQSSLELRISLLLIQSWEFPPVR 117

RESULT 8
Q14643 PREINITIARY: PRT: 202 AA
AC Q14643;
D: 01-JAN-1998 (TRMBLrel_05, Created)
DT 01-JAN-1998 (TRMBLrel_05, Last sequence update)
DT 01-JUN-2002 (TRMBLrel_21, Last annotation update)
DE Placental growth hormone 20kDa isoform precursor.
DE HGH-V.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-Term placenta;
RX MEDLINE-98373737; Pubmed-9709963;

```

ID	OT06015	PRELIMINARY:	PRT:	216 AA.
AC	070615:			
DT	01-AUG-1998	(T1EMBLrel. 07, Created)		
DT	01-AUG-1998	(T1EMBLrel. 07, Last sequence update)		
DT	01-MAR-2003	(T1EMBLrel. 23, Last annotation update)		
DE	Growth hormone precursor.			
OS	Spalax leucodon ehrenbergi (Ehrenberg's mole rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;			
OC	Nannospalax.			
OX	NCBI_taxid-30637;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE:99124645; PubMed:9924177;			
RA	Licupins A., Nevo E., Wallis M.;			
RT	*Cloning and characterisation of the gene encoding mole rat (Spalax			
RT	ehrenbergi) growth hormone.*			
RL	J. Mol. Endocrinol. 22:29-36(1999).			
DR	EMBL:	AJ005819; CAA06716.1;		
DR	HSP:	P0124; IAX;		
DR	InterPro:	IPRO01400; Somatotropin.		
DR	Tram:	PF00103; hormone; 1.		
DR	PRINTS:	PR00836; SOMATOTROPIN.		
DR	PROSITE:	PS00266; SOMATOTROPIN_1; 1.		
DR	PROSITE:	PS00336; SOMATOTROPIN_2; 1.		
DR	Signal:			
FT	SIGNAL:	1-26	POTENTIAL.	
FT	CHAIN:	27-216	GROWTH HORMONE.	
SO	SEQUENCE	216 AA: 24627 MW	FEA8A523HA0AUFF	CRC64;
Query Match: 65.2%; Score 306.5; DR 11; Length 216;				
Best Local Similarity 65.9%; Prod. No. 1.2e-26;				
Matches 60; Conservative 13; Mismatches 17; Indels 1; Gaps				
Qy	2	FTPIPLSRLEFNNMLRAHRLHGLAFTTYQEFREAIPEQKYSFIGNPUTSLSFSESPT 61		
Db	27	FPAMPLSNLFANAVLRAHRLHGLAADTYKEFRAY:PEQORYS-TONAAACFCSETIPA 85		
Qy	62	PSNRTEQKSKNLELRISLILQSMEPVO 92		
Db	86	PTKKEAQQKSWELLRSFLLIQLSWLGPVQ 116		
RESULT 11				
Q8MI73		PRELIMINARY:	PRT:	216 AA.
1D	Q8MI73:			
AC	Q8MI73:			
DT	01-OCT-2002	(T1EMBLrel. 22, Created)		
DT	01-OCT-2002	(T1EMBLrel. 22, Last sequence update)		
DT	01-MAR-2003	(T1EMBLrel. 23, Last annotation update)		
DE	Growth hormone precursor.			
GN	GH.			
OS	Delphinus delphis (Saddleback dolphin) (Black sea dolphin).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Delphinidae;			
OC	Delphinus.			
OX	NCBI_taxid-9728;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Manion Z., Wallis O.C., Wallis M.;			
RT	*Cloning and characterisation of the GH gene from the common dolphin			
RT	(Delphinus delphis)*.			
RL	Submitted (JUN-2002) to the EMBL/GenBank/DDRJ databases.			
DR	EMBL:	AJ492191; CAD37292.1;		
DR	InterPro:	IPR001400; Somatotropin.		
DR	Tram:	PF00103; hormone; 1.		
DR	PRINTS:	PR00836; SOMATOTROPIN.		
DR	PROSITE:	PS00266; SOMATOTROPIN_1; 1.		
DR	PROSITE:	PS00336; SOMATOTROPIN_2; 1.		
DR	Signal:			
FT	SIGNAL:	1-26	POTENTIAL.	
FT	CHAIN:	27-216	GROWTH HORMONE.	

```

DR Pfam: PF00103; hormone: 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
DR SO SEQUENCE 216 AA; 24682 MW; F72A06DA02536R18 CRC64;

Query Match 63.5%; Score 298.5; DB 11; Length 216;
Best Local Similarity 63.7%; Pred. No. 9,4e-26;
Matches 58; Conservative 14; Mismatches 18; Indels 1; Gaps 1;

QY 2 PPTPLSLRLFNAMLRRAHRLHQAFDTGYCFEAYIPKEQKYSFLONPOTSLSFSES IPT 61
DB 11 : ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
27 FPAPLSSLSFNNAVLRAGHLHQLAADITKKELRAYIPKQRYS-IQNQAACFCSESTIPA 85
QY 62 PSNREETOOKSNLELLRISTILLIQSWLPVQ 92
DB 11 : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
86 PTSKNEFAOGRIMELLRFSLILLIQSWLPVQ 114

RESULT 14
Q9JQM4
ID Q9JQM4 PRELIMINARY: PRI: 204 AA.
AC Q9JQM4;
DT 01-FEB-1997 (TRENBLrel_02, Created)
DI 01-FEB-1997 (TRENBLrel_02, Last sequence update)
DI 01-MAR-2003 (TRENBLrel_23, Last annotation update)
DE Placental growth hormone (Fragment).
OS Ovis aries (Sheep); Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Bovidae;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovioidea;
OC Bovidae; Caprinae; Ovis.
CN NCBI_TaxID=9940;
RX 11
RN KRC
RP SEQUENCE FROM N.A.
KC TISSUE:Placenta;
RX MEDLINE=97350833; PubMed=8895381;
RT Lactoxin M.C.; Deviney P.; Servely J.L.; Puissant C.; Kann G.;
RT *Expression of the growth hormone gene in ovine placenta: detection
RT and cellular localization of the protein.*;
RL Endocrinology 137:4886-4892(1996).
DR EMBL: U49063; AAC4879.1; -
DR HSP: P01241; IAXI
DR InterPro: IP001400; Somatotropin.
DR Pfam: PF00103; hormone: 1.
DK PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
FI NON_TER 1
SQ SEQUENCE 204 AA; 23462 MW; EBBB9451952635C5 CRC64;

Query Match 63.3%; Score 297.5; DB 6; Length 204;
Best Local Similarity 64.8%; Pred. No. 1.1e-25;
Matches 59; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

QY 2 PPTPLSLRLFNAMLRRAHRLHQAFDTGYCFEAYIPKEQKYSFLONPOTSLSFSES IPT 61
DB 11 : ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
15 FPANSLSLFPNAVLRAGHLHQLAADITKFERTYIPKQRYS-IQTGVACFCSESTIPA 73
QY 62 PSNREETOOKSNLELLRISTILLIQSWLPVQ 92
DB 11 : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
74 PTSKNEFAOGRIMELLRFSLILLIQSWLPVQ 104

RESULT 15
Q9JQM4
ID Q9JQM4 PRELIMINARY: PRI: 216 AA.
AC Q9JQM4;
DT 01-OCT-2000 (TRENBLrel_15, Created)
DI 01-OCT-2000 (TRENBLrel_15, Last sequence update)
DI 01-MAR-2003 (TRENBLrel_23, Last annotation update)
DE Growth hormone precursor.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

OC Mammalia: Eutheria: Rodentia: Hystricognathi: Caviidae: Cavia.  
 OX NCBI\_TaxID=10141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Odorico D.M., Fuller P.J., Herington A.C.,  
 RT "Cloning and sequence of guinea pig growth hormone (GH).";  
 RI Submitted (FEB-2000) to the EMBL/GenBank/ DDBJ databases.  
 DR EMBL: AF233853; AAF36409.1; -;  
 DR HSSP: P01241; IAXI.  
 DR InterPro: IPR001400; Somatotropin.  
 DR Pfam: PF00103; hormone; 1.  
 DR PRINTS: PR00836; SOMATOTROPIN.  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 KW Signal.  
 FT SIGNAL 1 26 POTENTIAL.  
 FT CHAIN 27 216 GROWTH HORMONE.  
 SQ SEQUENCE 216 AA: 24622 MW: 45500pt1; RQMRK QKGG;

Query Match 64.6% Score 29.1 E-16.111 Length 216.  
 Best Local Similarity 64.7% Posit. Pos. 1 26 27  
 Matches 58: Conservative 14: Mismatches 18: Indels 1: Gaps 1:  
 Oy 2 FVTPLSLRFDNMLRAHRLHOLAFDTYQEFEEAVIPKEQKYSFLONPQTSLSFSESIPT 61  
 II III II III III III III III III III III III III III III III  
 Db 27 FVAMPSSLSFGNVLRAOHLHOLAADTYKLRTYIPESQRYSLHNTQTAFCESETIPA 85  
 Oy 62 PSNREFTQKSNLELLRLSLILLIUSWLPVQ 92  
 III III III III III III III III III III III III III III III  
 Db 86 PTKKEFAQQRSDVELHFSLLIUSWLGVPQ 11

Search completed: September 15, 2003, 12:00:27  
 Job time : 35.9427 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:44:15 : Search time 60.4158 Seconds  
(without alignments)  
225,942 Million cell updates/sec

Title: US-09-423-100-4

Perfect score: 463

Sequence: 1 FVNHLCGSHLVNLYVCG.....IVEQCCTICSILYLENVGN 86

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1197863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

```

1: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1980.DAT.*
2: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1981.DAT.*
3: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1982.DAT.*
4: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1983.DAT.*
5: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1984.DAT.*
6: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1985.DAT.*
7: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1986.DAT.*
8: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1987.DAT.*
9: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1988.DAT.*
10: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1989.DAT.*
11: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1990.DAT.*
12: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1991.DAT.*
13: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1992.DAT.*
14: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1993.DAT.*
15: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1994.DAT.*
16: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1995.DAT.*
17: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1996.DAT.*
18: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1997.DAT.*
19: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1998.DAT.*
20: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA1999.DAT.*
21: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA2000.DAT.*
22: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA2001.DAT.*
23: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA2002.DAT.*
24: /SIDSI/qcdata/qeneseq/qeneseq-emb1/AA2003.DAT.*

```

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	463	100.0	86	5 AAP40829
2	463	100.0	86	17 AAR84061
3	463	100.0	86	20 AAY42858
4	463	100.0	86	21 AAB12770
5	463	100.0	86	23 AAM48218
6	463	100.0	87	3 AAP20036
7	463	100.0	87	5 AAP40217
8	463	100.0	87	6 AAP50127
9	463	100.0	87	6 AAP50060

10	463	100.0	87	7 AAP61090	Sequence encoded b
11	463	100.0	87	14 AAR32367	Proinsulin protein
12	463	100.0	88	11 AAR07682	Modified human ins
13	463	100.0	88	14 AAR33855	hpl. Homo sapiens
14	463	100.0	92	13 AAR20467	yeast alpha-factor
15	463	100.0	93	10 AAP90102	Synthetic proinsul
16	463	100.0	96	20 AAO17830	Human proinsulin #
17	463	100.0	96	20 AAY08004	Human proinsulin p
18	463	100.0	97	15 AAR68898	Human pro-insulin
19	463	100.0	97	21 AAB12773	Human proinsulin p
20	463	100.0	110	2 AAP10042	Sequence encoded b
21	463	100.0	110	2 AAP10053	Sequence of prepro
22	463	100.0	110	5 AAP40309	Sequence of human
23	463	100.0	110	20 AAY06608	Human preproinsuli
24	463	100.0	110	21 AAB26765	Human preproinsuli
25	463	100.0	110	21 AAB06144	Human insulin. Ho
26	463	100.0	110	21 AAY70366	Human Insulin used
27	463	100.0	110	21 AAY44367	Human proinsulin.
28	463	100.0	110	22 AAG55677	Human proinsulin n
29	463	100.0	110	22 AAB10337	Human pre proinsul
30	463	100.0	110	22 AAB5424	Secretory cell lin
31	463	100.0	110	23 AAG31590	Human preproinsuli
32	463	100.0	110	23 AAG60634	Human insulin. Ho
33	463	100.0	117	21 AAY69788	Human preproinsuli
34	463	100.0	130	21 AAY69787	MMWsp-MWmp10-Met-
35	463	100.0	51	20 AAW81856	MMWsp-MWmp10-(His
36	463	100.0	53	21 AAY53589	Human proinsulin a
37	463	100.0	51	20 AAW81857	Human preproinsuli
38	463	100.0	51	20 AAW81857	Human proinsulin a
39	463	100.0	88	14 AAR39574	Human Met-Arg proi
40	463	100.0	88	14 AAR39575	Human Met-Tyr proi
41	463	100.0	96	20 AAO17841	Human Met-phe proi
42	463	100.0	96	20 AAO17841	Human proinsulin #
43	463	100.0	97	16 AAR78661	Human proinsulin p
44	463	100.0	98	21 AAY51222	Proinsulin sequenc
45	463	100.0	99	21 AAY51223	Human preproinsuli

#### ALIGNMENTS

RESULT 1  
AAP40829  
ID AAP40829 standard; Protein; 86 AA.  
XX  
AC AAP40829;  
XX  
DI 25-MAR-2003 (updated)  
DI 03-AUG-1992 (first entry)  
XX  
DE Sequence of human insulin precursor.  
XX  
KW Insulin precursor; connecting peptide; diabetes; hormone.  
XX  
QS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Region 1..30 /label= chain H  
FT Peptide 31..65 /label= connecting peptide  
FT Region 66..86 /label= chain A  
FT Modified-site 1 /label= F-NH2-H  
FT /note= "H or a chemically or enzymatically cleavable AA residue or peptide residue"  
FT Disulfide bond 7..72  
FT Disulfide bond 19..85  
FT Disulfide bond 71..76  
FT Modified-site 86 /label= N-OH  
XX

PN US4430266-A.  
 XX 07-FEB-1984.  
 PD 16-FEB-1982; 82US-0349397.  
 XX 16-FEB-1982; 82US-0349397.  
 PR 27-MAR-1980; 80US-0134389.  
 PR 28-NOV-1980; 80US-0210696.  
 XX (E.L.H.) LILLY & CO. ELI.  
 PA Frank BH.  
 PI WPI: 1984-049032/08.  
 DR Insulin precursor prodn. from linear S-sulphonate and mercaptan  
 XX in single step without separate oxidn.  
 PT Claim 17; Column 4; 8pp; English.  
 XX The inventors claim a method for the prepn. of an insulin precursor  
 CC in which the A-chain and B-chain are joined through a connecting  
 CC peptide. The connecting peptide joins the A-chain at the amino  
 CC group of A-1 to the B-chain at the carboxyl group of B-30. The  
 CC method is pref. for the prepn. of human insulin precursor (see  
 CC AAP40829). The SEQs of the connecting peptides of a number of species  
 CC are given (see AAP40828, AAP40830-39).  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 XX Sequence 86 AA:  
 SQ Query Match 100.0%; Score 463; Db 5; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 8.4e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRRAEDLQVQLVGGGPGAGSLQPLALEG 60  
 DB 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRRAEDLQVQLVGGGPGAGSLQPLALEG 60  
 OY 61 SLQKRGIVEQCCTSCSLYQLENYCN 86  
 DB 61 SLQKRGIVEQCCTSCSLYQLENYCN 86  
 RESULT 2  
 ID AAR84061 standard; Protein; 86 AA.  
 XX AAR84061;  
 AC 22-AUG-1996 (first entry)  
 DT Human insulin.  
 DE Insulin; transformation; gene expression; fungi; fungal cell;  
 KW hormone; A-chain; C-chain; glycosylation.  
 KW Homo sapiens.  
 OS Key location/Qualifiers  
 XX CUS 1..261  
 FT /\*tag- a  
 FT /product- Insulin.  
 XX EP704527-A.  
 XX 03-APR-1996.  
 PD 03-AUG-1995; 95EP-0001999.  
 PF 05-AUG-1994; 94HR-0000432.  
 PR

PA (MEST/) MESTRIC S.  
 PA (PLIV ) PLIVA PHARM & CHEM FAH.  
 XX Mestric S, Punt PJ, Vallingar K, Van DEN HONDEL CAMJQ;  
 XX WPI: 1996-129917/14.  
 DR N-PSDB; AAT17830, AAT17831.  
 XX DNA encoding human insulin precursors - which comprise B- and  
 PT A-chains linked via amino acid chain contg. 1 or more glycosylation  
 PT sites, for prepn. of insulin in fungal cells  
 XX Disclosure: Figure 1; 32pp; English.  
 XX DNA sequences encoding insulin precursors of formula R-Pq A, where R  
 CC and A represent B- and A-chains of insulin respectively, and Pq  
 CC represents a modified C-peptide or any number of amino acids  
 CC comprising at least one glycosylation consensus site, can be  
 CC inserted into expression vectors which in turn can be used to  
 CC transform fungal host cells. The fungal cells are then cultured  
 CC and the insulin expressed in such cells can be harvested.  
 XX Sequence 86 AA:  
 SQ Query Match 100.0%; Score 463; Db 17; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 8.4e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRRAEDLQVQLVGGGPGAGSLQPLALEG 60  
 DB 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRRAEDLQVQLVGGGPGAGSLQPLALEG 60  
 OY 61 SLQKRGIVEQCCTSCSLYQLENYCN 86  
 DB 61 SLQKRGIVEQCCTSCSLYQLENYCN 86  
 RESULT 3  
 ID AAY42858 standard; Protein; 86 AA.  
 XX AAY42858;  
 AC 19-JAN-2000 (first entry)  
 DT Human insulin precursor, SEQ ID 5.  
 DE Insulin; precursor; growth hormone; chapetone; intramolecular;  
 KW folding; conformation; chimeric protein; clavabie; recombinant;  
 KW production; yield.  
 XX Homo sapiens.  
 OS WO9550302-A1.  
 XX 07-OCT 1999.  
 PD 31-MAR 1998; 98WO-CN00052.  
 PF 31-MAR-1998; 98WO-CN00052.  
 PR (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.  
 PA Gan Z;  
 PI WPI: 1999-610839/52.  
 DR New chimeric proteins containing human growth hormone fragment, used  
 PT particularly for the production of human insulin -  
 XX Claim 10; Page 29; 46pp; English.  
 XX This sequence represents a human insulin precursor comprising insulin



XX WPI: 2002-082942/11.  
 DR Treating cartilage disorders including cartilage damage by injury or  
 XX PT degenerative cartilaginous disorders, by contacting cartilage with  
 PT insulin-like growth factor analog with altered affinity for IGF-binding  
 PT proteins.  
 XX  
 PS Disclosure: Fig 16; 136pp; English.  
 XX  
 CC The present invention relates to a method for treating cartilage  
 CC disorders. The method comprises contacting cartilage with an active agent  
 CC such as insulin-like growth factor (IGF 1) analog with a binding affinity  
 CC preference for IGF binding protein-3 (IGFBP-3) over IGFBP-1, an IGF-1  
 CC analog with a binding affinity preference for IGFBP-1 over IGFBP-3, or a  
 CC IGFBP displacer peptide that prevents the interaction of IGF with an  
 CC IGFBP and does not bind to human IGF receptor. The method is useful for  
 CC treating cartilage disorders (CD), including degenerative CD, articular  
 CC CD such as rheumatoid arthritis and osteoarthritis. The present sequence  
 CC is human proinsulin, which was used to illustrate the invention.  
 XX  
 SQ Sequence 86 AA:

Query Match 100.0%; Score 463; DB 23; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 4e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRKRAEDLQVGQVELGGPGAGSLQPLALEG 60  
 DB 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRKRAEDLQVGQVELGGPGAGSLQPLALEG 60  
 QY 61 SLORRGIVEQCCTISGLYOLENYCN 86  
 DB 61 SLORRGIVEQCCTISGLYOLENYCN 86

RESULT 6  
 AAP20046  
 ID AAP20046 standard; Protein: 87 AA.  
 AC AAP20046;  
 XX  
 XX 25-MAR-2003 (updated)  
 DT 22-JUL-1992 (first entry)  
 DE Human proinsulin.  
 KW Proinsulin.  
 OS Homo sapiens.  
 XX EP55942-A.  
 XX  
 XX 14-JUL-1982.  
 PD 31 DEC-1981; 81EP-0306190.  
 PF 02-JAN-1981; 81US-0222019.  
 PR 23-JUL-1981; 81US-0286070.  
 PR 02-JAN-1982; 82US-0222010.  
 PR 03-MAR-1982; 82US-0354287.  
 XX (CNYNY-) STATE UNIV NEW YORK.  
 PA  
 PI Inoue M, Nakamura K;  
 XX WPI: 1982-59775E/29 (59775E).  
 DR N FSDB; AAN20041.  
 XX Plasmid cloning vehicles - useful for transforming bacterial  
 PT hosts to produce eukaryotic polypeptide(s).  
 XX  
 PS Disclosure: Fig 27; 114pp; English.

XX The sequence comprises human proinsulin.  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 XX  
 SQ Sequence 87 AA:  
 Query Match 100.0%; Score 463; DB 43; Length 87;  
 Best Local Similarity 100.0%; Pred. No. 4e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRKRAEDLQVGQVELGGPGAGSLQPLALEG 60  
 DB 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRKRAEDLQVGQVELGGPGAGSLQPLALEG 60  
 QY 61 SLORRGIVEQCCTISGLYOLENYCN 87  
 DB 61 SLORRGIVEQCCTISGLYOLENYCN 87  
 RESULT 7  
 AAP40217  
 ID AAP40217 standard; Protein: 87 AA.  
 AC AAP40217;  
 XX  
 XX 25-MAR-2003 (updated)  
 DT 12-FEB-1992 (first entry)  
 DE Sequence of the 32 N-terminal AAs of proinsulin.  
 KW Hormone; Cloning vector; phage resistant.  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FT Region 2...31  
 FT /label= B-chain  
 FT Region 32...66  
 FT /label= C-chain  
 FT Region 67...87  
 FT /label= A-chain  
 XX GB2126237-A.  
 PN 21-MAR-1984.  
 PD 01-SEP-1983; 83GB-0023468.  
 PF 03-SEP-1982; 82US-0414290.  
 PR 05-SEP-1984; 84US-0847338.  
 XX (EHL) LILLY & CO ELI.  
 PA  
 PI Hershester CL, Rostock PR;  
 XX WPI: 1984-070793/12.  
 DR N-FSDB; AAN40179.  
 XX Protecting bacteria from phage infection - by transformation with  
 PT cloning vector contg. segment with restriction and modification  
 PT activity  
 XX Example: Fig 10; 28pp; English.  
 PS  
 XX Plasmid pTH alpha 1 was constructed by inserting a synthesised gene  
 CC for thymosin alpha 1 (AAN40178) into plasmid pBR322. It is used for  
 CC the construction of pTRP24. The inventors claim a method for  
 CC protecting bacteria from phage infection - by transformation with  
 CC cloning vector contg. segment with restriction and modification  
 CC activity. Prodn. of plasmid pTR 26 or pTR27 which uses pTRP24; and  
 CC prodn. of plasmid pTR29 which uses a synthetic gene coding for the  
 CC 32 N-terminal AAs of proinsulin (see AAN40179).  
 CC (Updated on 25-MAR-2003 to correct PA field.)



CC readily isolated from cell lysates and is stable on washing with  
 CC urea or detergent solns. at low concns. The granule contains at  
 CC least 50% of proinsulin and all isolation operations are  
 CC simplified.

CC (Updated on 16-AUG-2002 to add missing OS field.)  
 CC (Updated on 25-MAR-2003 to correct PA field.)

XX  
 XX Sequence 87 AA;

Query Match 100.0%; Score 463; DB 6; Length 87;  
 Best Local Similarity 100.0%; Pred. No. 8.5e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRRAEDLQVQVELGGPGAGSLQPLALEG 60  
 DB 2 FVNHLCGSHLVEALYLVCGERGFFYTPKTRRAEDLQVQVELGGPGAGSLQPLALEG 61  
 OY 61 SLOKRGIVEOCTTSCSLYLENYCN 86  
 DB 62 SLOKRGIVEOCTTSCSLYLENYCN 87

RESULT 10  
 AAP61090  
 ID AAP61090 standard; Protein: 87 AA.

XX  
 XX AAK61090;

DT 28-FEB-1992 (first entry)

DE Sequence encoded by the structural gene for human proinsulin.

KW Recombinant plasmid; E.coli expression vector; secretion vector.

XX Homo sapiens.

PN US4624926-A.

XX  
 XX 25-Nov 1986.

XX  
 XX 03-MAR-1982; 82US-0354287.

PR 03-MAR-1982; 82US-0354287.

PR 02-JAN-1981; 81US-0222010.

PR 23-JUL-1981; 81US-0286070.

XX (UYNV) UNIV OF NEW YORK.

XX Inouye M, Nakamura K;

XX WPI: 1986 331802/50.

DR N-PSDB; AAN60872.

XX New recombinant plasmid(s) - contg. DNA sequences encoding  
 PT exogenous polypeptide and outer membrane protein of E.coli

XX Example: Fig 27; 44pp; English.

XX The inventors claim new recombinant plasmids contg. a DNA sequence  
 CC encoding a polypeptide, which is foreign to E.coli, in reading phase  
 CC with a DNA SQ, coding for at least one functional fragment derived  
 CC from an outer membrane lipoprotein gene of E.coli. The foreign gene  
 CC may be for human insulin. The lipoprotein gene functional fragment  
 CC may be the promoter, the 5'-UTR, the 3'-UTR or the transcription  
 CC termination signal provided that it includes at least the promoter.

XX Sequence 87 AA;

Query Match 100.0%; Score 463; DB 7; Length 87;  
 Best Local Similarity 100.0%; Pred. No. 8.5e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRRAEDLQVQVELGGPGAGSLQPLALEG 60

DB 2 FVNHLCGSHLVEALYLVCGERGFFYTPKTRRAEDLQVQVELGGPGAGSLQPLALEG 61  
 OY 61 SLOKRGIVEOCTTSCSLYLENYCN 86  
 DB 62 SLOKRGIVEOCTTSCSLYLENYCN 87

RESULT 11  
 AAR32367

ID AAR32367 standard; Protein: 87 AA.

XX AAR32367;

DT 25-MAR-2003 (updated)

DT 18-JUN-1993 (first entry)

XX Proinsulin protein sequence.

DE Human; proinsulin; vector; pGCI9; pPINS; CAT; pHC-CAT proinsulin;

XX Insulin analogue; type I; type II; diabetes.

XX Synthetic.

XX W05303174-A1.

XX 18-FEB-1993.

XX 31-JUL-1992; 92WO-0506451.

XX 08-AUG-1991; 91US-0741938.

XX 30-JUL-1992; 92US-0418453.

XX (PEIZ) PFIZER INC.

XX (SCIO) SCIOS INC.

XX Andy R.J. Larson Esq.

XX WPI: 1993-076530/09.

XX N-PSDB; AAO37003.

XX New hepato selective and peripheral selective human insulin

XX analogues - and their corresp. DNA, for treatment of type I and

XX type II diabetes

XX Disclosure: Fig 2b; 88pp; English.

XX This sequence represents human proinsulin and was decoded rom  
 CC the sequences given in AAO36996-7001. The cDNA fragment coding for  
 CC proinsulin was inserted into plasmid vector pUC19 and digested with  
 CC KpnI and HindIII. This resulted in the formation of the vector pPINS.  
 CC A fragment encoding amino acids 1-73 of CAT (see AAO37002) was inserted  
 CC into pPINS to give a plasmid which contained DNA sequences which coded  
 CC for amino acids 1-73 of CAT, an 8 amino acid linker sequence and human  
 CC proinsulin. This plasmid, pHC-CAT-proinsulin, could be used in the  
 CC formation of insulin analogues which may be used in the treatment of  
 CC types I and II diabetes.

XX (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 87 AA;

Query Match 100.0%; Score 463; DB 14; Length 87;  
 Best Local Similarity 100.0%; Pred. No. 8.5e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRRAEDLQVQVELGGPGAGSLQPLALEG 60  
 DB 2 FVNHLCGSHLVEALYLV ERGFFYTPKTRRAEDLQVQVELGGPGAGSLQPLALEG 61

OY 61 SLOKRGIVEOCTTSCSLYLENYCN 86

DB 62 SLOKRGIVEOCTTSCSLYLENYCN 87

Query Match	Rest Local Similarity	100.0%	Score 463	DR 11	Length 88
Matches	86	Conservative	0	Mismatches	0
Indels	0	Gaps	0		
QY	1	FVNHLCGSHLVLEALYLVCCGERGFFVTPKTRREADLQVGVELGGPGAGSLQPLALEG	60		
Db	3	FVNHLCGSHLVLEALYLVCCGERGFFVTPKTRREADLQVGVELGGPGAGSLQPLALEG	62		
QY	61	SLOKRGIVEQCCTSI CSLYLENYCN	86		
Db	63	SLOKRGIVEQCCTSI CSLYLENYCN	88		

DE Yeast alpha-factor signal-human proinsulin fusion product.  
 XX BCA-5; yeast preferred codons; post-translational processing;  
 KW insulin; endopeptidase.  
 XX Synthetic.

XX Key Location/Qualifiers  
 XX Cleavage site 6..7  
 FT /note= "signal proinsulin function"  
 FT Cleavage site 37..38

XX US5077204-A.

XX 31-DEC-1991.

XX 08-APR-1988; 88US-0183252.

XX 21-JUN-1984; 84US-0523306

XX 08-APR-1989; 87US-0183252

XX (REG ) UNIV CALIFORNIA.

XX Blake AJ, Blair LC, Julius D, Thornton W;

XX WPI; 1992-032671/04.

XX N-PSDB; AAQ20543.

XX Novel DNA for endopeptidase produ. - useful for in vivo or in  
 PT vitro processing of poly:peptide(s)

XX Example 1; Fig 1; 16pp; English.

XX The fusion product is encoded by a synthetic sequence having at its  
 CC 5'-end a modification of the 3'-end of the naturally occurring  
 CC alpha-factor secretory leader and processing signal sequence, where  
 CC three Gln Ala pairs have been deleted. A plasmid containing the  
 CC synthetic proinsulin coding sequence was used to transform kex2-  
 CC mutant yeast strains in the presence or absence of the cloned KEX2  
 CC gene. Post-translational processing of pro-insulin into peptides  
 CC only occurred in yeast transformed to kex2 plus. See also AAQ20545.  
 CC (updated on 25-MAR-2003 to correct PA field.)

XX Sequence 92 AA:

Query Match 100.0%; Score 463; DB 13; Length 92;  
 Best Local Similarity 100.0%; Pred. No. 90 43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNHLCGSHLVEALYLVGGERGFFYTPKTRREAEDLQVGVLEGGPGAGSLQPLALEG 60

Db 7 FVNHLCGSHLVEALYLVGGERGFFYTPKTRREAEDLQVGVLEGGPGAGSLQPLALEG 66

Oy 61 SLOKRGIVEQCCTSIICSLYLENYCN 86

Db 67 SLOKRGIVEQCCTSIICSLYLENYCN 92

RESULT 15

AAP90102

XX AAP90102 standard; protein; 93 AA.

XX AAP90102;

XX 25-MAR-2003 (updated)

DT 01-NOV-1989 (first entry)

XX Synthetic proinsulin.

XX Proinsulin; synthetic gene; yeast.

XX Homo sapiens.

XX Key Location/Qualifiers

FT 8..37

FT /note= "proinsulin chain B"

FT 38..39

FT /note= "dibasic processing site"

FT 38..72

FT /note= "proinsulin chain C"

FT 71..72

FT /note= "dibasic processing site"

FT 73..93

FT /note= "proinsulin chain A"

XX EP324274-A.

XX 19-JUL-1989.

XX 23-DEC-1988; 88EP-012306.

XX 30-DEC-1987; 87US-0196652.

XX (CHIR ) CHIRUKU CHIRU.

XX Tokapolski P;

XX WPI; 1989-208260/29.

XX N-PSDB; AAN90279.

XX Expression and secretion of heterologous proteins in yeast - using  
 PT truncated alpha-factor leader sequence contg. signal peptide and  
 PT glycosylation site.

XX Disclosure; fig 1; 24pp; English.

XX The proinsulin is expressed in *Saccharomyces cerevisiae* AB103.1.  
 CC It is encoded by a synthetic gene (see AAN90279). The 3 consecutive  
 CC peptides are proinsulin chains B, C and A resp. The regions are  
 CC dibasic processing sites.  
 CC (Updated on 25-MAR-2003 to correct P1 field.)

XX Sequence 93 AA:

Query Match 100.0%; Score 463; DB 10; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 410 43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNHLCGSHLVEALYLVGGERGFFYTPKTRREAEDLQVGVLEGGPGAGSLQPLALEG 60

Db 8 FVNHLCGSHLVEALYLVGGERGFFYTPKTRREAEDLQVGVLEGGPGAGSLQPLALEG 67

Oy 61 SLOKRGIVEQCCTSIICSLYLENYCN 86

Db 68 SLOKRGIVEQCCTSIICSLYLENYCN 94

Search completed: September 15, 2003, 12:00:58

Job time : 61.4158 secs



GenCore version 5.1.6  
Copyright (C) 1993 - 2003 Impugen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:57:40 : Search time 19.7276 Seconds  
(without alignments)  
184.449 Million cell updates/sec

Title: US-09-423-100-4

Perfect score: 463

Sequence: 1 F NOHLCGSHLVEALYLVCGR.....IVFOCTCSICSLYLENYCN 86

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 4210458 residues

Total number of hits satisfying the seq. for om-opts: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PT05\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARY:

Result No.	Score	Query Match	Length	DB ID	Description
1	463	100.0	86	US-09-477-924-2	Sequence 2, Appl
2	463	100.0	86	US-09-723-981-2	Sequence 2, Appl
3	463	100.0	86	US-09-723-846-2	Sequence 2, Appl
4	463	100.0	86	US-09-878-893-1	Sequence 1, Appl
5	463	100.0	96	US-09-134-1-4	Sequence 4, Appl
6	463	100.0	96	US-09-386-1A-4	Sequence 4, Appl
7	463	100.0	97	US-08-160-A-4	Sequence 4, Appl
8	463	100.0	110	US-08-950-A-11	Sequence 11, Appl
9	463	100.0	110	US-08-584-1-2	Sequence 2, Appl
10	463	100.0	110	US-08-784-1-2	Sequence 2, Appl
11	463	100.0	110	US-08-785-1-2	Sequence 2, Appl
12	463	100.0	110	US-08-472-1-1	Sequence 2, Appl
13	463	100.0	110	US-09-185-1-2	Sequence 2, Appl
14	463	100.0	110	PCT-US95-08596-2	Sequence 2, Appl
15	463	100.0	117	US-09-280-030-63	Sequence 63, Appl
16	463	100.0	130	US-09-280-030-62	Sequence 62, Appl
17	463	100.0	151	US-08-508-664-15	Sequence 15, Appl
18	463	100.0	161	US-08-508-664-16	Sequence 16, Appl
19	463	100.0	167	US-07-918-1-3	Sequence 8, Appl
20	463	100.0	167	US-08-081-1-8	Sequence 8, Appl
21	457	98.7	96	US-09-134-1-5	Sequence 5, Appl
22	457	98.7	96	US-09-386-1A-5	Sequence 5, Appl
23	457	98.7	97	US-08-389-4-7-7	Sequence 7, Appl
24	456	98.5	90	US-08-030-731A-43	Sequence 43, Appl
25	446	96.3	97	US-09-099-1-7-6	Sequence 6, Appl
26	444	95.9	97	US-09-099-307-8	Sequence 8, Appl
27	443	95.7	110	US-08-589-028-4	Sequence 4, Appl

28	443	95.7	110	3	US-08-784-582-4	Sequence 4, Appl
29	443	95.7	110	3	US-08-785-271-4	Sequence 4, Appl
30	440	95.0	97	3	US-09-099-307-7	Sequence 7, Appl
31	435	94.0	97	3	US-09-099-307-11	Sequence 11, Appl
32	398	86.0	91	4	US-09-676-787-7	Sequence 7, Appl
33	292.5	63.2	67	4	US-08-981-988A-1	Sequence 1, Appl
34	240.5	62.7	83	4	US-08-981-988A-3	Sequence 3, Appl
35	288.5	62.3	83	4	US-08-981-988A-6	Sequence 6, Appl
36	288.5	62.3	113	4	US-09-484-848-16	Sequence 16, Appl
37	286.5	61.9	67	4	US-08-981-988A-2	Sequence 2, Appl
38	285.5	61.7	83	4	US-08-981-988A-8	Sequence 8, Appl
39	284.5	61.4	67	4	US-08-981-988A-5	Sequence 5, Appl
40	281.5	60.8	67	4	US-08-981-988A-7	Sequence 7, Appl
41	281	60.7	86	4	US-09-201-227A-43	Sequence 43, Appl
42	280	60.5	112	4	US-09-484-848-15	Sequence 15, Appl
43	279.5	60.4	83	4	US-08-981-988A-9	Sequence 9, Appl
44	278.5	60.2	69	4	US-08-981-988A-4	Sequence 4, Appl
45	272.5	58.9	53	1	US-08-211-617-4	Sequence 4, Appl

#### ALIGNMENTS

RESULT 1  
US-09-477-924-2  
: Sequence 2, Application: US/09477924  
: Patent No. 6403764  
: GENERAL INFORMATION:  
: APPLICANT: Dubaqui, Yves  
: APPLICANT: Lomman, Henry  
: TITLE OF INVENTION: PROTEIN VARIANTS  
: FILE REFERENCE: P1712R1-1  
: CURRENT APPLICATION NUMBER: US/09/477,924  
: CURRENT FILING DATE: 2000-11-05  
: NUMBER OF SEQ ID NOS: 6  
: SEQ ID NO 2  
: LENGTH: 86  
: TYPE: PRT  
: ORGANISM: Homo sapiens  
US-09-477-924-2

Query Match 100.0% Score 463, DB 4; Length 86;  
Best Local Similarity 100.0% Pred. No. 1.5e-47;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 FVNHLCGSHLVEALYLVCGRGFYTPKTRRAEFLQNGQVGLGGPGAGSLQPLALEG 60  
DB 1 FVNHLCGSHLVEALYLVCGRGFYTPKTRRAEFLQNGQVGLGGPGAGSLQPLALEG 60  
QY 6: SLQKRGIVEQCTCSICSLYLENYCN 86  
DB 61 SLQKRGIVEQCTCSICSLYLENYCN 86

RESULT 2  
US-09-723-981-2  
: Sequence 2, Application US/09723981  
: Patent No. 6506874  
: GENERAL INFORMATION:  
: APPLICANT: Dubaqui, Yves  
: APPLICANT: Lomman, Henry  
: TITLE OF INVENTION: PROTEIN VARIANTS  
: FILE REFERENCE: P1712R1  
: CURRENT APPLICATION NUMBER: US/09/723,981  
: CURRENT FILING DATE: 2000-11-28  
: PRIOR APPLICATION NUMBER: 09/477,923  
: PRIOR FILING DATE: 2000-01-05  
: NUMBER OF SEQ ID NOS: 6  
: SEQ ID NO 2  
: LENGTH: 86  
: TYPE: PRT  
: ORGANISM: Homo sapiens  
US-09-723-981-2

Query Match 100.0% Score 463; DB 4; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-47;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYVCGERGFFYTPKIRREAEQLQGVQVELGGGPGAGSLOPLALEG 60  
 DB 1 FVNHLCGSHLVEALYVCGERGFFYTPKIRREAEQLQGVQVELGGGPGAGSLOPLALEG 60

QY 61 SLOKRGIVEQCCTSCISLYOLENYCN 86  
 DB 61 SLOKRGIVEQCCTSCISLYOLENYCN 86

## RESULT 3

US-09-723-896 2  
 : Sequence 2, Application US/09723896  
 : Patent No. 6504443  
 : GENERAL INFORMATION:  
 : APPLICANT: Dabagile, Yves  
 : APPLICANT: Loman, Henry  
 : TITLE OF INVENTION: PROTEIN VARIANTS  
 : FILE REFERENCE: P1712R1  
 : CURRENT APPLICATION NUMBER: US/09/723,896  
 : CURRENT FILING DATE: 2000-11-28  
 : PRIOR APPLICATION NUMBER: US/09/477,923  
 : PRIOR FILING DATE: 2000-01-05  
 : NUMBER OF SEQ ID NOS: 6  
 : SEQ ID NO: 2  
 : LENGTH: 86  
 : TYPE: PRT  
 : ORGANISM: Homo sapiens  
 US-09-723-896 2

Query Match 100.0% Score 463; DB 4; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-47;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYVCGERGFFYTPKIRREAEQLQGVQVELGGGPGAGSLOPLALEG 60  
 DB 1 FVNHLCGSHLVEALYVCGERGFFYTPKIRREAEQLQGVQVELGGGPGAGSLOPLALEG 60

QY 61 SLOKRGIVEQCCTSCISLYOLENYCN 86  
 DB 61 SLOKRGIVEQCCTSCISLYOLENYCN 86

## RESULT 4

US-09-878-380 1  
 : Sequence 1, Application US/09878380  
 : Patent No. 6534281  
 : GENERAL INFORMATION:  
 : APPLICANT: Fujire Bio Inc.  
 : APPLICANT: KITAJIMA, Sachiko  
 : APPLICANT: KURAMO, Yoshihiro  
 : APPLICANT: NAKATSUBO, Kaoru  
 : APPLICANT: NISHIZONO, Isao  
 : TITLE OF INVENTION: Immunoassay For Measuring Human C-Peptide and Kit Therefor  
 : FILE REFERENCE: 0760-0291P  
 : CURRENT APPLICATION NUMBER: US/09/878,380  
 : CURRENT FILING DATE: 2001-06-12  
 : PRIOR APPLICATION NUMBER: JP 2000-174669  
 : PRIOR FILING DATE: 2000-06-12  
 : NUMBER OF SEQ ID NOS: 2  
 : SOFTWARE: PatentIn version 3.1  
 : SEQ ID NO: 1  
 : LENGTH: 86  
 : TYPE: PRT  
 : ORGANISM: Homo sapiens  
 US-09-878-380-1

Query Match 100.0% Score 463; DB 4; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-47;

Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYVCGERGFFYTPKIRREAEQLQGVQVELGGGPGAGSLOPLALEG 60  
 DB 1 FVNHLCGSHLVEALYVCGERGFFYTPKIRREAEQLQGVQVELGGGPGAGSLOPLALEG 60

QY 61 SLOKRGIVEQCCTSCISLYOLENYCN 86  
 DB 61 SLOKRGIVEQCCTSCISLYOLENYCN 86

## RESULT 5

US-09-134-836 4  
 : Sequence 4, Application US/09134836  
 : Patent No. 5986048  
 : GENERAL INFORMATION:  
 : APPLICANT: Kubroder, Franz Josef  
 : APPLICANT: Keller, Reinhold  
 : TITLE OF INVENTION: Improved process for obtaining  
 : TITLE OF INVENTION: Resulting precursors having currently bonded cystine bridges  
 : NUMBER OF SEQUENCES: 7  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Finnegan, Henderson, Farabee, Garrett &  
 : STREET: 1300 I Street, N.W.  
 : CITY: Washington  
 : STATE: D.C.  
 : COUNTRY: USA  
 : ZIP: 20005-3415  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: floppy disk  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/09/134,836  
 : FILING DATE:  
 : CLASSIFICATION:  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Leslie McDowell  
 : REGISTRATION NUMBER: 34,872  
 : REFERENCE/DOCKET NUMBER: 02481-1600-00000  
 : TELEPHONE: (202) 408-4000  
 : TELEFAX: (202) 408-4400  
 : INFORMATION FOR SEQ ID NO: 4:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 96 amino acids  
 : TYPE: amino acid  
 : STRANDEDNESS: single  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: protein  
 : ORIGINAL SOURCE:  
 : ORGANISM: Escherichia coli  
 : FEATURE:  
 : NAME/KEY: Protein  
 : LOCATION: 1..96  
 : US-09-134-836-4

Query Match 100.0% Score 463; DB 4; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-47;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYVCGERGFFYTPKIRREAEQLQGVQVELGGGPGAGSLOPLALEG 60  
 DB 1 FVNHLCGSHLVEALYVCGERGFFYTPKIRREAEQLQGVQVELGGGPGAGSLOPLALEG 70

QY 61 SLOKRGIVEQCCTSCISLYOLENYCN 86  
 DB 71 SLOKRGIVEQCCTSCISLYOLENYCN 96

## RESULT 6

```

US-09-386-303A-4
: Sequence 4, Application US/09386303A
: Patent No. 6380355
: GENERAL INFORMATION:
: APPLICANT: Rudroder, Franz-Josef
: TITLE OF INVENTION: Improved process for obtaining
: insulin precursors having correctly bonded cystine bridges
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farrelow, Garrett &
: Dunner
: STREET: 1300 I Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC DOS 5/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/386,303A
: FILING DATE: 31-Aug-1999
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 09/114,816
: FILING DATE: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: Leslie McDonell
: REGISTRATION NUMBER: 34,872
: REFERENCE/DOCKET NUMBER: 02481 1600-00000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 408-4000
: TELEFAX: (202) 408-4400
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 96 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Escherichia coli
: FEATURE:
: NAME/KEY: Protein
: LOCATION: 1..96
: SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-386-303A-4
Query Match 100.0%; Score 463; DB 4; Length 96;
Best Local Similarity 100.0%; Pred. No. 1.8e-47;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 FVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVGAGSLQPLALEG 60
Db 11 FVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVGAGSLQPLALEG 70
QY 61 SLKRGIVEQCCTSCSLYQLENYCN 86
Db 71 SLKRGIVEQCCTSCSLYQLENYCN 96
RESULT 7
US-08-160-376A-4
: Sequence 4, Application US/08160376A
: Patent No. 5473049
: GENERAL INFORMATION:
: APPLICANT: Obermeier, Ranier
: APPLICANT: Gerli, Martin
: APPLICANT: Ludwig, Jurgen
: APPLICANT: Sabel, Walter

```

```

: TITLE OF INVENTION: Process For Obtaining Proinsulin
: TITLE OF INVENTION: Possessing Correctly Linked
: TITLE OF INVENTION: Cystine Bridges
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Kenneth A. Genoni, Esq.
: STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500
: CITY: Somerville
: STATE: New Jersey
: COUNTRY: U.S.A.
: ZIP: 08876-1258
: COMPUTER READABLE FORM:
: MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
: COMPUTER: IBM 386
: OPERATING SYSTEM: WINDOWS 3.1
: SOFTWARE: WORDPERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/160,376A
: FILING DATE: December 1, 1993
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GE P 424,420.7
: FILING DATE: December 2, 1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Barbara V. Maurel, Esq.
: REGISTRATION NUMBER: 31,287
: REFERENCE/DOCKET NUMBER: HGE 92/F 384
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (908) 231-4079
: TELEFAX: (908) 231-2255
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 97 Amino Acids
: TYPE: Amino Acid (AA)
: TOPOLOGY: not relevant
US-08-160-376A-4
Query Match 100.0%; Score 463; DB 1; Length 97;
Best Local Similarity 100.0%; Pred. No. 1.8e-47;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 FVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVGAGSLQPLALEG 60
Db 12 FVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVGAGSLQPLALEG 71
QY 61 SLKRGIVEQCCTSCSLYQLENYCN 86
Db 72 SLKRGIVEQCCTSCSLYQLENYCN 97
RESULT 8
US-08-950-720A-11
: Sequence 11, Application US/08950720A
: Patent No. 6046028
: GENERAL INFORMATION:
: APPLICANT: Conklin, Darrell C.
: APPLICANT: Lottin-Day, Catherine E.
: APPLICANT: Lok, Si
: APPLICANT: Jaspers, Stephen R.
: TITLE OF INVENTION: INSULIN HOMOLOG
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ZymoGenetics, Inc.
: STREET: 1201 Eastlake Avenue East
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98102
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSEQ for Windows Version 2.0

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/950,720A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sawislak, Deborah A
REGISTRATION NUMBER: 37,438
REFERENCE/DOCKET NUMBER: 96-09
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6672
TELEFAX: 206-442-6678
TELEX:
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 110 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 60486280
US-08-950-720A 11

```

```

Query Match      100.0%: Score 464; DB 3; Length 110;
Best Local Similarity 100.0%: Pred. No. 2,1e-47;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEIQQVGQVGLGGPGAGSLQPLALEG 60
DB 25 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEIQQVGQVGLGGPGAGSLQPLALEG 84
QY 61 SLQKRGIVEOCCTSGCSLYOLENYCN 86
DB 85 SLQKRGIVEOCCTSGCSLYOLENYCN 110

```

```

RESULT 9
US-08-589-028-2
Sequence 2, Application US/08589028
Patent No. 6087129
GENERAL INFORMATION:
APPLICANT: Newgard, Christopher B.
APPLICANT: Halban, Philippe A.
APPLICANT: No. 6087129minington, Karl D.
APPLICANT: Clark, Samuel A.
APPLICANT: Thiipen, Anice E.
APPLICANT: Quade, Christian
APPLICANT: Kruse, Fred
TITLE OF INVENTION: Recombinant Expression of Proteins From
TITLE OF INVENTION: SECRETORY CELL LINES
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/589,028
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 47,642
REFERENCE/DOCKET NUMBER: 47,642
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/416-3000
TELEFAX: 512/474-7577

```

```

TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 110 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-589-028-2

Query Match      100.0%: Score 463; DB 3; Length 110;
Best Local Similarity 100.0%: Pred. No. 2,1e-47;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEIQQVGQVGLGGPGAGSLQPLALEG 60
DB 25 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEIQQVGQVGLGGPGAGSLQPLALEG 84
QY 61 SLQKRGIVEOCCTSGCSLYOLENYCN 86
DB 85 SLQKRGIVEOCCTSGCSLYOLENYCN 110

```

```

RESULT 10
US-08-784-582-2
Sequence 2, Application US/08784582
Patent No. 6110707
GENERAL INFORMATION:
APPLICANT: Newgard, Christopher B.
APPLICANT: Halban, Philippe A.
APPLICANT: No. 6110707minington, Karl D.
APPLICANT: Clark, Samuel A.
APPLICANT: Thiipen, Anice E.
APPLICANT: Quade, Christian
APPLICANT: Kruse, Fred
APPLICANT: McGarry, Dennis
TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM
TITLE OF INVENTION: SECRETORY CELL LINES
NUMBER OF SEQUENCES: 79
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/784,582
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,427
FILING DATE: 15-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/589,028
FILING DATE: 19-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: UTSD:514
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/416-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 110 amino acids
TYPE: amino acid
STRANDEDNESS:

```



```

; CURRENT FILING DATE: 1998-11-04
; EARLIER APPLICATION NUMBER: 60/087,660
; EARLIER FILING DATE: 1998-06-02
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-185-852-2
Query Match      100.0%   Score 463; DB 4; Length 110;
Best Local Similarity 100.0%; Pred. No. 2,1e-47;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 60
DDB 25 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 64
QY 61 SLOKRGIIVEQCCTSIICSLYLENYCN 86
DDB 85 SLOKRGIIVEQCCTSIICSLYLENYCN 117

RESULT 14
PCT-US95-08596-2
Sequence 2: Application PC/TUS9508596
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Proinsulin Peptide compounds for detecting
NUMBER OF SEQUENCES: 23 diabetes
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/08596
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/272,220
FILING DATE: 08-JULY-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Bonatti, Giulio A., Jr.
REGISTRATION NUMBER: 41,503
REFERENCE/PACKET INFORMATION: DCI 092PC
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 110 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-08596-2

```

```

Query Match      100.0%  Score 463  I/B 5  Length 110;
Best Local Similarity 100.0%  Pred. No. 2,1e-47;
Matches      86;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY      1  FVNHLGCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQVGLGGSPGAGSLQPLALEG 60

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein : protein search, using sw model

Run on: September 15, 2003, 12:03:35 : Search time 36.0645 seconds  
(without alignments)  
347.945 Million cell updates/sec

Title: US-09-423-100-4

Perfect score: 463

Sequence: 1 FVNHLCGSHLVYALVLCG.....IVEQCCTSGSLVYLENYCN 86

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 541936

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA:

```
1: /cgn2_6/ptodata/1/pubpa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpa/US09A_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpa/US10H_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	463	100.0	86	10	US-09-878-380-1
2	463	100.0	86	11	US-09-858-935B-4
3	463	100.0	86	14	US-10-028-113-2
4	463	100.0	86	14	US-10-054-873-4
5	463	100.0	96	10	US-09-947-663-4
6	463	100.0	110	9	US-09-295-8-125
7	463	100.0	110	9	US-09-815-229-3
8	463	100.0	110	10	US-09-804-409A-9
9	463	100.0	110	12	US-09-969-748C-6
10	463	100.0	110	15	US-10-038-686-1
11	463	100.0	110	15	US-10-328-813-2
12	463	100.0	117	9	US-09-280-030-63
13	463	100.0	130	9	US-09-280-030-62
14	457	98.7	96	10	US-09-947-563-5
15	306	66.1	166	9	US-09-925-297-805

16	300	64.8	56	9	US-09-815-229-10	Sequence 10, Appl
17	285	61.6	54	9	US-09-815-229-13	Sequence 13, Appl
18	267	57.7	52	14	US-10-054-873-5	Sequence 5, Appl
19	267	57.7	107	14	US-10-054-873-6	Sequence 6, Appl
20	267	57.7	150	14	US-10-054-873-7	Sequence 7, Appl
21	261.5	56.5	51	11	US-09-858-935B-5	Sequence 5, Appl
22	261.5	56.5	51	14	US-10-028-410-3	Sequence 3, Appl
23	258.5	55.8	124	10	US-09-894-711-18	Sequence 18, Appl
24	258.5	55.8	138	10	US-09-861-687-39	Sequence 19, Appl
25	256	55.3	58	14	US-10-066-009A-3	Sequence 3, Appl
26	250.5	54.1	124	9	US-09-736-611-12	Sequence 12, Appl
27	250.5	54.1	125	9	US-09-736-611-10	Sequence 10, Appl
28	250.5	54.1	147	9	US-09-736-611-8	Sequence 8, Appl
29	249.5	53.9	124	9	US-09-740-359-12	Sequence 12, Appl
30	249.5	53.9	124	10	US-09-894-711-12	Sequence 12, Appl
31	249.5	53.9	125	9	US-09-740-359-10	Sequence 10, Appl
32	249.5	53.9	125	10	US-09-894-711-10	Sequence 10, Appl
33	249.5	53.9	147	9	US-09-740-359-7	Sequence 7, Appl
34	248	53.6	144	9	US-09-740-359-5	Sequence 5, Appl
35	248	53.6	144	9	US-09-740-359-5	Sequence 5, Appl
36	248	53.6	146	10	US-09-894-711-5	Sequence 5, Appl
37	227.5	49.1	144	10	US-09-894-711-7	Sequence 132, App
38	209	45.3	46	9	US-09-205-658-134	Sequence 133, App
39	197	42.5	46	9	US-09-205-658-133	Sequence 115, App
40	184.5	39.8	50	9	US-09-205-658-115	Sequence 135, App
41	180	38.9	46	9	US-09-205-658-135	Sequence 14, Appl
42	178	38.4	42	9	US-09-815-229-14	Sequence 6, Appl
43	178	38.4	32	10	US-09-947-563-6	Sequence 134, App
44	175	37.8	46	9	US-09-205-658-134	Sequence 4, Appl
45	174	37.6	35	9	US-09-815-229-4	

#### ALIGNMENTS

RESULT 1  
US-09-878-380-1  
Sequence 1: Application US/09878380  
Patent No: US20020160435A1  
GENERAL INFORMATION:  
APPLICANT: Fujirebio Inc.  
APPLICANT: KITAJIMA, Sachiko  
APPLICANT: KURANO, Yoshihiro  
APPLICANT: NAKATSURU, Kaoru  
APPLICANT: NISHIZONO, Isao  
TITLE OF INVENTION: Immunoassay For Measuring Human C-Peptide and Kit Therefor  
FILE REFERENCE: 0760-0291P  
CURRENT APPLICATION NUMBER: US/09/878,380  
PRIOR FILING DATE: 2001-06-12  
PRIOR APPLICATION NUMBER: JP 2000-174691  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 86  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-878-380-1

Query Match 100.0% Score 443; DB 10; Length 86;  
Best Local Similarity 100.0% Pred. No. 9.7e 45;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNHLCGSHLVYALVLCGGEFFYTPKTRAEADLVGVGVGGAGSLQIALEG 60  
|||||  
Db 1 FVNHLCGSHLVYALVLCGGEFFYTPKTRAEADLVGVGVGGAGSLQIALEG 60  
Qy 61 SLOKRGIVEQCCTSGSLVYLENYCN 86  
|||||  
Db 61 SLOKRGIVEQCCTSGSLVYLENYCN 86

RESULT 2

```

US-09-858-945B-4
: Sequence 4, Application US/09858935B
: Publication No. US20030069177A1
: GENERAL INFORMATION:
: APPLICANT: Dubaque, Yves
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Lowman, Henry B.
: TITLE OF INVENTION: METHOD FOR TREATING CARTILAGE DISORDERS
: FILE REFERENCE: P1794R1
: CURRENT APPLICATION NUMBER: US/09/858,935B
: PRIOR FILING DATE: 2002-07-02
: PRIOR APPLICATION NUMBER: US 60/248,985
: PRIOR FILING DATE: 2000-11-15
: PRIOR APPLICATION NUMBER: US 60/204,450
: PRIOR FILING DATE: 2000-05-16
: NUMBER OF SEQ ID NOS: 153
: SEQ ID NO 4
: LENGTH: 86
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-858-945B-4

Query Match 100.0% Score 453 DB 14 Length 86
Best Local Similarity 100.0% Pred. No. 9.7e-45
Matches 86 Conservative 0 Mismatches 0 Indels 0 Gaps 0

QY 1 FVNHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGVEFGGPGAGSLQPLALEG 60
DB 1 FVNHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGVEFGGPGAGSLQPLALEG 60

QY 61 SLKRGIVQCCTSIQSLYLENYCN 86
DB 61 SLKRGIVQCCTSIQSLYLENYCN 86

RESULT 4
US-10-028-410-2
: Sequence 2, Application US/10028410
: Publication No. US20020160955A1
: GENERAL INFORMATION:
: APPLICANT: Dubaque, Yves
: APPLICANT: Lowman, Henry
: TITLE OF INVENTION: PROTEIN VARIANTS
: FILE REFERENCE: P1712R1-1
: CURRENT APPLICATION NUMBER: US/10/028,410
: CURRENT FILING DATE: 2001-12-19
: PRIOR APPLICATION NUMBER: US/09/473,324
: PRIOR FILING DATE: 2000-01-05
: NUMBER OF SEQ ID NOS: 6
: SEQ ID NO 2
: LENGTH: 86
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-028-410-2

Query Match 100.0% Score 453 DB 14 Length 86
Best Local Similarity 100.0% Pred. No. 9.7e-45
Matches 86 Conservative 0 Mismatches 0 Indels 0 Gaps 0

QY 1 FVNHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGVEFGGPGAGSLQPLALEG 60
DB 1 FVNHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGVEFGGPGAGSLQPLALEG 60

QY 61 SLKRGIVQCCTSIQSLYLENYCN 86
DB 61 SLKRGIVQCCTSIQSLYLENYCN 86

RESULT 5
US-09-947-563-4
: Sequence 4, Application US/09947563
: Patent No. US20020156234A1
: GENERAL INFORMATION:
: APPLICANT: Rubbreker, Franz-Josef
: APPLICANT: Keller, Reinhold
: TITLE OF INVENTION: Improved process for obtaining
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS: insulin precursors having correctly bonded cysteine bridge
: ADDRESSSEE: Finnegan, Henderson, Farrahaw, Garrett &
: Dunner
: STREET: 1300 I Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible

```



OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION NUMBER: US/09/947,563  
FILING DATE: 07-Sep-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/134,836  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Leslie McDonnell  
REGISTRATION NUMBER: 34,872  
REFERENCE/DOCKET NUMBER: 02481.16.00 00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 96 amino acids  
TYPE: amino acid  
SUSANBUENESS, Stuart  
LOCALITY: Lincoln  
MOLECULAR TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..96  
SEQUENCE DESCRIPTION: SEQ ID NO: 4  
US-09-947-563-4

Query Match 100.0%; Score 463; DB 10; Length 96;  
Best Local Similarity 100.0%; Pred. No. 1.3e-44;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 60  
|||||  
DB 11 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 70  
|||||

QY 61 SLOKRGIVEQCCTSCSLYQLENYCN 86  
|||||  
DB 71 SLOKRGIVEQCCTSCSLYQLENYCN 96  
|||||

RESULT 6  
US-09-205-658-125  
Sequence 125, Application US/09205658  
Patent No. US20010029617A1  
GENERAL INFORMATION:  
APPLICANT: Ravkun, Gaty  
APPLICANT: Csq, Scott  
TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR  
TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS  
FILE REFERENCE: 00786/351004  
CURRENT APPLICATION NUMBER: US/09/205,658  
CURRENT FILING DATE: 1998-12-03  
EARLIER APPLICATION NUMBER: 08/857,076  
EARLIER FILING DATE: 1997-05-15  
EARLIER APPLICATION NUMBER: 08/888,514  
EARLIER FILING DATE: 1997-07-07  
EARLIER APPLICATION NUMBER: US98/10080  
EARLIER FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 328  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 125  
LENGTH: 110  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-205-658-125

Query Match 100.0%; Score 463; DB 9; Length 110;  
Best Local Similarity 100.0%; Pred. No. 1.3e-44;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 60  
|||||  
DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 84  
|||||

QY 61 SLOKRGIVEQCCTSCSLYQLENYCN 86  
|||||  
DB 85 SLOKRGIVEQCCTSCSLYQLENYCN 110  
|||||

RESULT 7  
US-09-815-229-3  
Sequence 3, Application US/09815229  
Patent No. US20020058614A1  
GENERAL INFORMATION:  
APPLICANT: Filvaroff, Ellen H.  
APPLICANT: Okumu, Franklin W.  
TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGENOUS DISORDERS  
FILE REFERENCE: P1786RIUS  
CURRENT APPLICATION NUMBER: US/09/815,229  
CURRENT FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: US 62/192,173  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 17  
SEQ ID NO 3  
LENGTH: 110  
TYPE: PRT  
ORGANISM: Homo sap ius  
US-09-815-229-3

Query Match 100.0%; Score 463; DB 9; Length 110;  
Best Local Similarity 100.0%; Pred. No. 1.3e-44;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 60  
|||||  
DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 84  
|||||

QY 61 SLOKRGIVEQCCTSCSLYQLENYCN 86  
|||||  
DB 85 SLOKRGIVEQCCTSCSLYQLENYCN 110  
|||||

RESULT 8  
US-09-804-409A-9  
Sequence 9, Application US/09804409A  
Patent No. US20020155100A1  
GENERAL INFORMATION:  
APPLICANT: KIEFER, TIMOTHY J.  
APPLICANT: CHEUNG, ANTHONY T.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATED PROTEIN  
TITLE OF INVENTION: EXPRESSION IN GUT  
FILE REFERENCE: 029996/027 8721  
CURRENT APPLICATION NUMBER: US/09/804,409A  
CURRENT FILING DATE: 2001-03-12  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 9  
LENGTH: 110  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-804-409A-9

Query Match 100.0%; Score 463; DB 10; Length 110;  
Best Local Similarity 100.0%; Pred. No. 1.3e-44;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 60  
|||||  
DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 84  
|||||

QY 61 SLOKRGIVEQCCTSCSLYQLENYCN 86  
|||||

Db 85 SLOKRGIVEQCCTSIICSLYOLENYCN 110

## RESULT 9

US-09-969-748C-6  
Sequence 6, Application US/09969748C  
Publication No. US20030161809A1

## GENERAL INFORMATION:

APPLICANT: ARIZERE PHARMACEUTICALS, INC.  
APPLICANT: HOUSTON, Lou, L.  
APPLICANT: SHERIDAN, Philip, J.  
APPLICANT: HAWLEY, Stephen  
APPLICANT: GLYNN, Jacqueline, M.  
APPLICANT: CHAPIN, Steven  
APPLICANT: BASU, Amresh

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TRANSPORT OF BIOLOGICALLY ACTIVE AGENTS ACROSS CELLULAR BARRIERS

FILE REFERENCE: 057220-0303

CURRENT APPLICATION NUMBER: US 60/969,748C

PRIOR FILING DATE: 2002-12-10

PRIOR APPLICATION NUMBER: US 60/267,493

PRIOR FILING DATE: 2003-02-24

PRIOR APPLICATION NUMBER: US 60/245,819

PRIOR FILING DATE: 2000-11-14

PRIOR APPLICATION NUMBER: US 60/248,478

PRIOR FILING DATE: 2000-11-13

PRIOR APPLICATION NUMBER: US 60/237,929

PRIOR FILING DATE: 2000-10-02

NUMBER OF SEQ ID NOS: 115

SOFTWARE: PatentIn version 3.0

SEQ ID NO: 6

LENGTH: 110

TYPE: PRT

ORGANISM: Homo sapiens

US-09-969-748C-6

## Query Match

Best Local Similarity 100.0%; Score 463; DB 12; Length 110;

Mismatches 86; Conservative 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYVGCGERGFYTPKTRREAEDLVQGVLEGGPGAGSLOPLALEG 60

DB 25 FVNOHLCGSHLVEALYVGCGERGFYTPKTRREAEDLVQGVLEGGPGAGSLOPLALEG 84

QY 61 SLOKRGIVEQCCTSIICSLYOLENYCN 86

DB 85 SLOKRGIVEQCCTSIICSLYOLENYCN 110

## RESULT 10

US-10-038-686-1  
Sequence 1, Application US/10038686  
Publication No. US20030045467A1

## GENERAL INFORMATION:

APPLICANT: Orban, Tibamir  
TITLE OF INVENTION: AUTOANTIGEN VACCINE

FILE REFERENCE: 10276-067001

CURRENT APPLICATION NUMBER: US/10/038,686

PRIOR FILING DATE: 2002-05-31

PRIOR APPLICATION NUMBER: 60/266,068

PRIOR FILING DATE: 2001-01-05

NUMBER OF SEQ ID NOS: 6

SOFTWARE: FastSeq for Windows Version: 4.0

SEQ ID NO: 1

LENGTH: 110

TYPE: PRT

ORGANISM: Homo sapiens

US-10-038-686-1

## Query Match

Best Local Similarity 100.0%; Score 463; DB 15; Length 110;

Mismatches 86; Conservative 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYVGCGERGFYTPKTRREAEDLVQGVLEGGPGAGSLOPLALEG 60  
DB 25 FVNOHLCGSHLVEALYVGCGERGFYTPKTRREAEDLVQGVLEGGPGAGSLOPLALEG 84

QY 61 SLOKRGIVEQCCTSIICSLYOLENYCN 86

DB 85 SLOKRGIVEQCCTSIICSLYOLENYCN 110

## RESULT 11

US-10-328-813-2

Sequence 2, Application US/10328813

Publication No. US20030113305A1

## GENERAL INFORMATION:

APPLICANT: Osbourne, William R. A.

APPLICANT: Ramesh, Nagarajan

TITLE OF INVENTION: Compositions and Methods for Treating Diabetes

FILE REFERENCE: P-W 3264

CURRENT APPLICATION NUMBER: US/10/328,813

PRIOR FILING DATE: 2002-12-24

PRIOR APPLICATION NUMBER: US/09/785,452

PRIOR FILING DATE: 1998-11-04

PRIOR APPLICATION NUMBER: 60/887,666

NUMBER OF SEQ ID NOS: 11

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO: 2

LENGTH: 110

TYPE: PRT

ORGANISM: Homo sapiens

US-10-328-813-2

Query Match 100.0%; Score 454; DB 15; Length 110;

Best Local Similarity 100.0%; Pred. No. 1.3e-44;

Mismatches 86; Conservative 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYVGCGERGFYTPKTRREAEDLVQGVLEGGPGAGSLOPLALEG 60

DB 25 FVNOHLCGSHLVEALYVGCGERGFYTPKTRREAEDLVQGVLEGGPGAGSLOPLALEG 84

QY 61 SLOKRGIVEQCCTSIICSLYOLENYCN 86

DB 85 SLOKRGIVEQCCTSIICSLYOLENYCN 110

## RESULT 12

US-09-280-030-63

Sequence 63, Application US/09280030A

Patent No. US20010021515A1

## GENERAL INFORMATION:

APPLICANT: Sato, Seiji

APPLICANT: Higashikuni, Naohiko

APPLICANT: Kudo, Toshiyuki

APPLICANT: Kondo, Masaki

TITLE OF INVENTION: DNAS ENCODING NEW FUSION PROTEINS AND PROCESSES FOR

TITLE OF INVENTION: PREPARING USEFUL POLYPEPTIDES THROUGH EXPRESSION OF THE

FILE REFERENCE: 382 1026

CURRENT APPLICATION NUMBER: US/09/280,030A

CURRENT FILING DATE: 1999-03-26

EARLIER APPLICATION NUMBER: JP10-87339/1998

EARLIER FILING DATE: 1998-03-31

NUMBER OF SEQ ID NOS: 66

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO: 63

LENGTH: 117

TYPE: PRT

ORGANISM: Artificial: Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Designated is

OTHER INFORMATION: an amino acid sequence of

OTHER INFORMATION: MW:sp-MWmp10-Met-Proinsulin

US-09-280-030-63

Query Match 100.0% Score 463 DB 9 Length 117;  
 Best local Similarity 100.0% Pred. No. 1.4e-44;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVGGERGFFYTPKTRAEADLVQGVLEGGPGAGSLQPLALEG 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 32 FVNHLCGSHLVEALYLVGGERGFFYTPKTRAEADLVQGVLEGGPGAGSLQPLALEG 91  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 61 SLQKRGIVEQCCTSCSLYLENYCN 86  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 92 SLQKRGIVEQCCTSCSLYLENYCN 117  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 13  
 US-09-280-030 62  
 Sequence 62 Application US/09280030A  
 Patent No. US20016021515A1  
 GENERAL INFORMATION:  
 APPLICANT: Saito, Seiji  
 APPLICANT: Blazhshikuni, Naoki-ko  
 APPLICANT: Kudo, Toshiyuki  
 APPLICANT: Kudo, Masaki  
 TITLE OF INVENTION: DNAs ENCODING NEW HLS-ON PROTEINS AND PROCESSES FOR  
 TITLE OF INVENTION: PREPARING USEFUL POLYPEPTIDES THROUGH EXPRESSION OF THE  
 TITLE OF INVENTION: DNAs  
 FILE REFERENCE: 382.1026  
 CURRENT APPLICATION NUMBER: US/09/280,0  
 CURRENT FILING DATE: 1999-03-26  
 EARLIER APPLICATION NUMBER: JP10-87319/  
 EARLIER FILING DATE: 1998-03-11  
 NUMBER OF SEQ ID NOS: 66  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 62  
 LENGTH: 130  
 TYPE: PRT  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: Description of Artificial Sequence: Designated is  
 OTHER INFORMATION: an amino acid sequence of  
 OTHER INFORMATION: MWPsp-MWmp10-(His)6-Linker-Met-Proinsulin  
 US-09-280-030 62

Query Match 100.0% Score 463 DB 9 Length 130;  
 Best local Similarity 100.0% Pred. No. 1.6e-44;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVGGERGFFYTPKTRAEADLVQGVLEGGPGAGSLQPLALEG 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 45 FVNHLCGSHLVEALYLVGGERGFFYTPKTRAEADLVQGVLEGGPGAGSLQPLALEG 104  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 61 SLQKRGIVEQCCTSCSLYLENYCN 86  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 105 SLQKRGIVEQCCTSCSLYLENYCN 130  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 14  
 US-09-947-563-5  
 Sequence 5 Application US/09947563  
 Patent No. US2002156234A1  
 GENERAL INFORMATION:  
 APPLICANT: Rubroder, Franz-Josef  
 APPLICANT: Keller, Reinhold  
 TITLE OF INVENTION: Improved process for obtaining  
 TITLE OF INVENTION: insulin precursors having correctly bonded cysteine bridges  
 NUMBER OF SEQUENCES: 7  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
 Dunner  
 STREET: 1300 I Street, N.W.  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA

ZIP: 20005-3715  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/947,563  
 FILING DATE: 07-Sep-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 09/134,836  
 FILING DATE: <Unknown>  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Leslie McDonell  
 REGISTRATION NUMBER: 34,872  
 REFERENCE/DOCKET NUMBER: 62481.1600.00306  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 408-4000  
 TELEFAX: (202) 408-4400  
 INTERNET E-MAIL FOR SEQ ID NOS: 51  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 94 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLUCULE TYPE: protein  
 ORIGIN: NCBI SOURCE:  
 ORGANISM: Escherichia coli  
 FEATURE:  
 NAME/KEY: Protein  
 LOCATION: 1..96  
 SEQUENCE DESCRIPTION: SEQ ID NO: 51  
 US-09-947-563-5

Query Match 94.7% Score 457 DB 10 Length 96;  
 Best local Similarity 100.0% Pred. No. 5.2e-44;  
 Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVGGERGFFYTPKTRAEADLVQGVLEGGPGAGSLQPLALEG 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 11 FVNHLCGSHLVEALYLVGGERGFFYTPKTRAEADLVQGVLEGGPGAGSLQPLALEG 70  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 61 SLQKRGIVEQCCTSCSLYLENYCN 85  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 71 SLQKRGIVEQCCTSCSLYLENYCN 95  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 15  
 US-09-925-297-805  
 Sequence 805 Application US/09925297  
 Patent No. US20020081659A1  
 GENERAL INFORMATION:  
 APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
 FILE REFERENCE: PA105  
 CURRENT APPLICATION NUMBER: US/09/925,297  
 CURRENT FILING DATE: 2001-08-10  
 PRIOR APPLICATION NUMBER: PCT/US00/05989  
 PRIOR FILING DATE: 2000-03-08  
 PRIOR APPLICATION NUMBER: 60/124,270  
 PRIOR FILING DATE: 1999-03-12  
 NUMBER OF SEQ ID NOS: 428  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 805  
 LENGTH: 166  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: SITE  
 LOCATION: (84)  
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 NAME/KEY: SITE

```
; LOCATION: (92)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (105)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (124)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (132)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (144)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (145)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (165)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-423-297-805
```

```
Query Match: 56.1%; Score: 390; 48 to 1 length 176.
Best Local Similarity: 96.7%; Pred. No. 5; 2e-27;
Matches: 58; Conservative: 0; Mismatch: 3; 2; Indels: 0; Gaps: 0;
Oy 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTTPAEDLQVGQVELGGGPGAGSIQPLALEG 60
Db 42 FVNHLCGSHLVEALYLVCGERGFFYTPKTTPAEDLQVGQVELGGGPGAGSIQPLALEG 101
```

```
Search completed: September 15, 2003, 12:23:35
Job time : 36.0645 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

GM protein - protein search, using SW model

Run on: September 15, 2003, 11:56:45 / Search time 14.1792 seconds  
(without alignments)  
583.284 Million cell updates/sec

Title: US-09-423-100-4

Perfect score: 453

Sequence: 1 FVNHLCGSHAEVLYVGG.....IVPGQCTGICSLYGLNTN 86

Scoring table: BL2SUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 95168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB Seq Length: 0

Maximum DB Seq Length: 200000000

Post-processing: Minimum Match: 0%

Maximum Match: 100%

Listing first 45 summaries

Database: PIR76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	463	100.0	110	1	IPNG
2	461	100.0	110	2	A42179
3	456	98.5	110	2	H42179
4	456	98.5	110	2	J00178
5	424	91.6	110	3	INR6
6	417	90.1	110	3	IPG5
7	414	85.1	86	3	IPH0
8	404	85.1	110	1	IPK72
9	404	85.1	110	1	INNS2
10	392	84.7	108	2	A39883
11	392	84.7	110	2	I48166
12	385	83.2	110	1	IPK1
13	383	82.7	84	1	IPG3
14	366.5	79.2	105	1	IPB9
15	346	79.0	108	1	INNS1
16	334.5	72.2	108	2	S0227H
17	320.5	69.2	77	1	INSH
18	314	67.8	110	1	IP3P
19	277.5	59.9	109	1	IPKTHU
20	276.5	59.7	103	2	I51221
21	265.5	57.3	106	1	IPXL2
22	265.5	57.3	107	1	IPCH
23	262.5	56.7	106	1	IPXL1
24	256.5	55.4	51	1	INWHP
25	256.5	55.4	51	1	INWHF
26	256.5	55.4	51	1	INEL
27	256.5	55.4	81	1	IPK
28	256	55.3	96	2	PC7082
29	254.5	55.0	51	1	INHY

30	251.5	54.3	51	1	INMSP
31	250.5	54.1	51	2	A59151
32	246.5	53.2	51	1	INWHL5
33	246.5	53.2	51	1	INOT
34	246.5	53.2	51	2	INCM4
35	245.5	53.0	51	1	INCI
36	244.5	52.8	51	1	INMKSU
37	239.5	51.7	51	2	J00362
38	234.5	50.6	51	1	INCR
39	231.5	50.0	51	1	INGS
40	227.5	49.1	51	1	INPU
41	227.5	49.1	51	1	INTK
42	227.5	49.1	51	1	INOS
43	227.5	49.1	51	1	A61149
44	227.5	49.1	51	2	A60414
45	225	48.6	52	2	S44459

## ALIGNMENTS

### RESULT 1

IPNG

Insulin precursor [arabidopsis] - human

N. Alterman, Robert, precursor insulin

C. Species: Homo sapiens (man)

C. Date: 24-Oct-1981 #Sequence\_revision 23 Oct 1981 #TextLchange 08-Dec-2000

C. Accession: A93222; A94253; A94216; A94251; A93144; A92075; A9186; 158114; A01579

R. Bell, G. L. Pictet, R. L. Rutter, W. J. Cordell, B. Tischer, E. Goodman, H. M.

Nature 284, 26-32, 1980

A. Title: Sequence of the human insulin gene.

A. Reference number: A93222; M01180120725; PMID:6244748

A. Accession: A93222

A. Molecule type: DNA

A. Residues: 1-110 <REL>

A. Cross-references: GR:J00265; NID:q186429; PIDN:AAA59172.1; PID:q36828

R. Ullrich, A. Dull, F. J. Gray, A. Brosius, J. Sures, I.

Science 209, 612-615, 1980

A. Title: Genetic variation in the human insulin gene.

A. Reference number: A94253; M01080236313; PMID:6248962

A. Accession: A94253

A. Molecule type: DNA

A. Residues: 1-110 <REL>

A. Cross-references: GR:J00265; NID:q186429; PIDN:AAA59172.1; PID:q36828

R. Bell, G. L. Pictet, R. L. Rutter, W. J. Cordell, B. Tischer, E. Goodman, H. M., J.

Nature 282, 525-527, 1979

A. Title: Nucleotide sequence of a cDNA clone encoding human preproinsulin.

A. Reference number: A93216; M01080054779; PMID:503234

A. Accession: A93216

A. Molecule type: mRNA

A. Residues: 1-110 <REL2>

A. Cross-references: GR:J00265; NID:q186429; PIDN:AAA59172.1; PID:q36828

R. Sures, I. Ullrich, D. V. Gray, A. Ullrich, A.

Science 208, 57-59, 1980

A. Title: Nucleotide sequence of human preproinsulin complementary DNA.

A. Reference number: A94251; M01080147417; PMID:6927840

A. Accession: A94251

A. Molecule type: mRNA

A. Residues: 1-110 <REL>

A. Cross-references: GR:J00265; NID:q186429; PIDN:AAA59172.1; PID:q36828

R. Nicol, D. S. H. W. Smith, L. F.

Nature 187, 483-485, 1960

A. Title: Amino acid sequence of human insulin.

A. Reference number: A93144

A. Accession: A93144

A. Molecule type: protein

A. Residues: 25-54; 90-110 <NIC>

R. Oyer, P. E. Cho, S. Peterson, J. D. Steiner, D. F.

J. Biol. Chem. 246, 1375-1386, 1971

A. Title: Studies on human proinsulin. Isolation and amino acid sequence of the huma

A. Reference number: A92075; M0107116410; PMID:5101771

A. Accession: A92075

A. Molecule type: protein

A:Residues: 57-87 <OYE>  
 R:Ko, A.; Smyth, D.G.; Markussen, J.; Sundby, F.  
 Eur. J. Biochem. 20, 190-199, 1971  
 A:Title: Amino acid sequence of the C-peptide of human proinsulin.  
 A:Reference number: A91186; MUID:71257722; PMID:5560404  
 A:Accession: A91186  
 A:Molecule type: protein  
 A:Residues: 57-87 <KOA>  
 R:Lucassen, A.M.; Julier, C.; Beressi, J.P.; Moitard, C.; Froguet, M.; Bell  
 Nature Genet. 4, 305-310, 1993  
 A:Title: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment  
 A:Reference number: I58114; MUID:93364428; PMID:8358440  
 A:Accession: I58114  
 A:Status: preliminary; translated from GR/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-59,63-110 <RFS>  
 A:Cross-references: GR:L15440; NID:q307071; PIDN:AA59179.1; PID:q307072  
 R:Sieber, P.; Kamber, B.; Hartmann, A.; Joehli, A.; Rinkler, B.; Ritte, W.  
 Helv. Chim. Acta 57, 2617-2621, 1974  
 A:Title: Totalsynthese von Humaninsulin unter azobischer Bildung der Disulfidbindung.  
 A:Reference number: A91636; MUID:75077273; PMID:4444294  
 A:Contents: annotation; synthesis  
 A:Note: disulfide bonded human insulin was synthesized. The synthesized hormone was identical  
 R:Naithani, V.K.  
 Hoppe-Seyler's Z. Physiol. Chem. 354, 659-672, 1973  
 A:Title: The synthesis of C-peptide of human proinsulin.  
 A:Reference number: A91658; MUID:75040007; PMID:4803904  
 A:Contents: annotation; synthesis of residues 57-87  
 R:Geiger, R.; Jaeger, G.; Koenig, W.  
 Chem. Ber. 106, 2347-2352, 1973  
 A:Title: Synthesis of the complete sequence of human proinsulin C-peptide and its [Glu-9  
 A:Reference number: A90914  
 A:Contents: annotation; synthesis of residue 57-87  
 R:Kaufmann, J.E.; Irminger, J.C.; Halban, P.A.  
 Biochem. J. 110, 869-874, 1995  
 A:Title: Sequence requirements for proinsulin processing at the B-chain/C peptide junction  
 A:Reference number: S58661; MUID:96013185; PMID:7575420  
 A:Contents: annotation; site-directed mutagenesis study of proteolytic processing  
 C:Genetics:  
 A:Gene: GDB:INS  
 A:Cross references: GDB:119349; OMIM:176730  
 A:Map position: 11p15.3-11p15.5  
 A:Introns: 63/1  
 C:Superfamily: insulin  
 C:Keywords: hormone; pancreas  
 F:1-24/Domain: signal sequence \*status predicted <SIG>  
 F:25-54/Domain: insulin chain B \*status experimental <RGH>  
 F:55-90-110/Product: insulin \*status experimental <MAT>  
 F:57-87/Domain: connecting C peptide \*status experimental <CEP>  
 F:90-110/Domain: insulin chain A \*status experimental <ACH>  
 F:31-96,44-104,95-100/Disulfide bonds: \*status experimental

Query Match 100.0%; Score 463; DB 1; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLGGPGAGSLOPLALEG 60  
 DB 25 FVNHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLGGPGAGSLOPLALEG 84

QY 61 SLOKRGIVQCCTSTCSLYOLENYCN 86  
 DB 85 SLOKRGIVQCCTSTCSLYOLENYCN 110

RESULT 2  
 A42179  
 Insulin precursor - chimpanzee  
 C:Species: Pan troglodytes (chimpanzee)  
 C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 16-Jul-1999  
 C:Accession: A42179; S22058  
 R:Seino, S.; Bell, G.I.; Li, W.H.

Mol. Biol. Evol. 9, 193-203, 1992  
 A:Title: Sequences of primate insulin genes support the hypothesis of a slower rate  
 A:Reference number: A42179; MUID:92219533; PMID:1560757  
 A:Accession: A42179  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-110 <SEI>  
 A:Cross-references: EMBL:X61089; NID:q48251; PIDN:CAA43403.1; PID:q48252  
 A:Note: sequence extracted from NCBI backbone (NCBI:95067)  
 C:Genetics:  
 A:Introns: 63/2  
 C:Superfamily: insulin

Query Match 100.0%; Score 463; DB 2; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLGGPGAGSLOPLALEG 60  
 DB 25 FVNHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLGGPGAGSLOPLALEG 84

QY 61 SLOKRGIVQCCTSTCSLYOLENYCN 46  
 DB 85 SLOKRGIVQCCTSTCSLYOLENYCN 110

RESULT 3  
 A42179  
 Insulin precursor - green monkey  
 C:Species: Cercopithecus aethiops (green monkey, aivet)  
 C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 16-Jul-1999  
 A:Accession: A42179; A05232; S:6494; S22056  
 R:Seino, S.; Bell, G.I.; Li, W.H.  
 Mol. Biol. Evol. 9, 193-203, 1992  
 A:Title: Sequences of primate insulin genes support the hypothesis of a slower rate  
 A:Reference number: A42179; MUID:92219533; PMID:1560757  
 A:Accession: A42179  
 A:Molecule type: DNA  
 A:Residues: 1-110 <SEI>  
 A:Cross-references: EMBL:X61092; NID:q22808; PIDN:CAA43405.1; PID:q22809  
 A:Note: sequence extracted from NCBI backbone (NCBI:95185; NCBI:95194)  
 R:Peterson, J.D.; Nehrllich, S.; Oyer, P.E.; Steiner, D.F.  
 J. Biol. Chem. 247, 4866-4871, 1972  
 A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog proin-

A:Reference number: A92133; MUID:7258016; PMID:4626469  
 A:Accession: A05232  
 A:Molecule type: protein  
 A:Residues: 57-87 <PEI>  
 C:Genetics:  
 A:Introns: 63/1  
 C:Superfamily: insulin  
 C:Keywords: hormone; pancreas  
 F:1-24/Domain: signal sequence \*status predicted <SIG>  
 F:25-54/Domain: insulin chain B \*status predicted <RGH>  
 F:55-90-110/Product: insulin \*status predicted <MAT>  
 F:57-87/Domain: connecting peptide \*status experimental <CEP>  
 F:90-110/Domain: insulin chain A \*status predicted <ACH>  
 F:31-96,43-109,95-100/Disulfide bonds: \*status predicted

Query Match 98.5%; Score 456; DB 2; Length 110;  
 Best Local Similarity 98.8%; Pred. No. 1.6e-42;  
 Matches 85; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLGGPGAGSLOPLALEG 60  
 DB 25 FVNHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLGGPGAGSLOPLALEG 84

QY 61 SLOKRGIVQCCTSTCSLYOLENYCN 86  
 DB 85 SLOKRGIVQCCTSTCSLYOLENYCN 110

RESULT 4

QY   61 SLQKRGIVQCCTSTCSLYOLENYCN 86  
     :|||||:|||||:|||||:  
Db   85 ALQKRGI VEQCCTSTCSLYOLENYCN 110

RESULT 6

IPOG

insulin precursor - dog

C:Species: Canis lupus familiaris (dog)  
C>Date: 24-Apr-1984 #sequence\_revision 15-Nov-1984 #text\_change 16-Jul-1999  
C:Accession: A92413; A01587; S16493  
R:Kwok, S.C.M.; Chan, S.J.; Steiner, D.F.  
J. Biol. Chem. 258, 2357-2363, 1983  
A:Title: Cloning and nucleotide sequence analysis of the dog insulin gene. Coded am  
A:Reference number: A92413; MUID:83109071; PMID:6296142  
A:Accession: A92413  
A:Molecule type: protein  
A:Residues: 1-110 <SMIT>  
A:Cross-references: GB:V00179; GR:100042; NID:q941; PIDN:CAA23475.1; PID:g995  
R:Smith, I.F.  
Am. J. Med. 40, 662-666, 1966  
A:Title: Species variation in the amino acid sequence of insulin.  
A:Reference number: A90029; MUID:66160119; PMID:5449593  
A:Accession: A01587  
A:Molecule type: Protein  
A:Residues: 25-54;90-110 <SMIT>  
R:Petersen, J.D.; Nefflich, S.; Oyer, P.E.; Steiner, D.F.  
J. Biol. Chem. 247, 4866-4871, 1972  
A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog pro  
A:Reference number: A92111; MUID:72258016; PMID:4626369  
A:Accession: S16493  
A:Molecule type: protein  
A:Residues: 55-85;1\*,8? <PEI>  
C:Superfamily: insulin  
C:Keywords: hormone; pancreas  
F:1-24/Domain: signal; sequence #status predicted <SIG>  
F:25-54/Domain: insulin chain B #status experimental <BCH>  
F:55-54, 90-110/Product: insulin #status experimental <MAT>  
F:57-87/Domain: connecting peptide #status predicted <CPDP>  
F:90-110/Domain: insulin chain A #status experimental <ACH>  
F:31-96,43-109,95-100/Disulfide bonds: #status experimental

Query Match             90.1%      Score 417; dB 1; Length 120;  
Best Local Similarity   89.5%      Pred. No. 2.9e-38;  
Matches                  77; Conservative   1; Mismatches   8; Indels   0; Gaps   0;

QY   1 FYNHLCGGSHLVEALYLVCGERGFFYTPKTREARDAIQGVQLVGCGSPGCAGSIOPALEG 60  
Db   25 FYNHLCGGSHLVEALYLVCGERGFFYTPKTREARDELDLVGRIVEELQVRNDELAVGAPEGGLPALEG 84

QY   61 SLQKRGIVQCCTSTCSLYOLENYCN 86  
     :|||||:|||||:|||||:  
Db   85 ALQKRGI VEQCCTSTCSLYOLENYCN 110

RESULT 7

IPOG

insulin precursor - horse

C:Species: Equus caballus (domestic horse)  
C>Date: 13-Jul-1981 #sequence\_revision 13-Jul-1981 #text\_change 16-Jul-1999  
C:Accession: A01580; A92120  
R:Harris, J.I.; Sanger, F.; Naughton, M.A.  
Arch. Biochem. Biophys. 65, 427-428, 1956  
A:Title: Species differences in insulin.  
A:Reference number: A90382  
A:Accession: A01580  
A:Molecule type: protein  
A:Residues: 1-30;66-86 <HAR>  
R:Tager, H.S.; Steiner, D.F.  
J. Biol. Chem. 247, 7936-7940, 1972  
A:Title: Primary structures of the proinsulin connecting peptides of the rat and ho  
A:Reference number: A92120; MUID:73061498; PMID:4440931

A:Accession: A92120  
A:Molecule type: protein  
A:Residues: 33-63 <TAG>  
C:Comment: X's at positions 31-32 and 64-65 represent paired basic residues assumed (by  
C:Superfamily: insulin  
F:1-30/Domain: insulin chain B #status experimental <BCH>  
F:1-30,66-86/Product: insulin #status experimental <MAT>  
F:33-63/Domain: connecting peptide #status experimental <CPP>  
F:66-86/Domain: insulin chain A #status experimental <ACH>  
F:7-72,75-85,91-96/Disulfide bonds: #status predicted

Query Match 85.1%; Score 394; DB 1; Length 86;  
Best Local Similarity 84.9%; Pred. No. 7.2e-36;  
Matches 73; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVYALVLCGERGFFYTPKTRAEADLQVGVGVGGPGAGSLQPLALEG 60  
DB 1 FVNOHLCGSHLVYALVLCGERGFFYTPKTRAEADLQVGVGVGGPGAGSLQPLALEG 60  
QY 61 SLQKRGIVQCCTSCISLYOLENYCN 86  
DB 61 SLQKRGIVQCCTSCISLYOLENYCN 86

RESULT 8  
IPRT2  
insulin 2 precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 23-Mar-1981 #sequence,revision: 23-Oct-1981 #text\_change 24-Sep-1999  
C:Accession: B90789; B94231; C92120; I64880; A01590; B92120  
R:Lomedico, P.; Rosenthal, N.; Efstratiadis, A.; Gilbert, W.; Kolodner, R.; Tizard, R.  
Cell 18, 545-558, 1979  
A:Title: The structure and evolution of the two nonallelic rat preproinsulin genes.  
A:Reference number: A90789; MUID:80045035; PMID:458284  
A:Accession: B90789  
A:Molecule type: DNA  
A:Residues: 1-110 <LON>  
A:Cross references: GB:J00748; NID:q204958; PIDN:AAA41443.1; PID:q204959  
R:Steiner, D.F.; Clark, J.L.; Nolan, C.; Rabinstein, A.H.; Margolis, E.; Allen, B.; Oye  
Recent Prot. Form. Res. 25, 207-282, 1969  
A:Title: Proinsulin and the biosynthesis of insulin.  
A:Reference number: A94231; MUID:70067613; PMID:431193H  
A:Accession: B94231  
A:Molecule type: protein  
A:Residues: 25-54;90-110 <STE>  
R:Tager, H.S.; Steiner, D.F.  
J. Biol. Chem. 247, 7936-7940, 1972  
A:Title: Primary structures of the proinsulin connecting peptides of the rat and horse.  
A:Reference number: A92120; MUID:73061498; PMID:4640931  
A:Accession: C92120  
A:Molecule type: protein  
A:Residues: 57-87 <TAG>  
R:Lomedico, P.T.; Rosenthal, N.; Kolodner, R.; Efstratiadis, A.; Gilbert, W.  
Ann. N. Y. Acad. Sci. 343, 425-432, 1980  
A:Title: The structure of rat preproinsulin gene.  
A:Reference number: 151945; MUID:80240379; PM 249167  
A:Accession: I64880  
A:Status: preliminary; translated from GB/EMBL/UDBI  
A:Molecule type: DNA  
A:Residues: 1-110 <RES>  
A:Cross references: GB:M25585; NID:q204950; PIDN:AAA41440.1; PID:q204952  
C:Genetics:  
A:Gene: INS2  
A:Introns: 63/1  
C:Superfamily: insulin  
C:Keywords: hormone; pancreas  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-54/Domain: insulin chain B #status experimental <BCH>  
F:57-87/Domain: connecting peptide #status experimental <MAT>  
F:90-110/Domain: insulin chain A #status experimental <ACH>  
F:31-96,43-109,95-100/Disulfide bonds: #status experimental

Query Match 85.1%; Score 394; DB 1; Length 110;  
Best Local Similarity 84.9%; Pred. No. 9.2e-36;  
Matches 73; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVYALVLCGERGFFYTPKTRAEADLQVGVGVGGPGAGSLQPLALEG 60  
DB 25 FVNOHLCGSHLVYALVLCGERGFFYTPKTRAEADLQVGVGVGGPGAGSLQPLALEG 84  
QY 61 SLQKRGIVQCCTSCISLYOLENYCN 86  
DB 85 ARQKRGIVQCCTSCISLYOLENYCN 110

RESULT 9  
INMS2  
insulin 2 precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 31-Mar-1992 #sequence,revision: 14-Jul-1994 #text\_change 18-Jun-1999  
C:Accession: A26342; B46172; A61012; B01592  
R:Wentworth, H.M.; Schaefer, J.M.; Villa-Komaroff, L.; Chirgwin, J.M.  
J. Mol. Biol. 23, 305-312, 1986  
A:Title: Characterization of the two nonallelic genes encoding mouse preproinsulin.  
A:Reference number: A92565; MUID:87169768; PMID:414503  
A:Accession: A26342  
A:Molecule type: DNA  
A:Residues: 1-110 <MEN>  
A:Cross references: GB:XG4724; NID:q52714; PIDN:CAA28433.1; PID:q52715  
R:Sawa, T.; Ohgaki, S.; Morioaka, H.; Yano, S.  
J. Mol. Endocrinol. 5, 61-67, 1990  
A:Title: Molecular cloning and DNA sequence analysis of preproinsulin genes in the N  
A:Reference number: A48172; MUID:90372989; PMID:2397023  
A:Accession: B48172  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-110 <SAW>  
R:Linde, S.; Nielsen, J.H.; Hansen, B.; Wellinder, R.S.  
J. Chromatogr. 462, 243-254, 1989  
A:Title: Reversed-phase high-performance liquid chromatographic analyses of insulin  
A:Reference number: A61012; MUID:89292078; PMID:2661585  
A:Accession: A61012  
A:Molecule type: protein  
A:Residues: 57-87 <LIN>  
R:Buenzli, H.F.; Glatthaar, B.; Kunz, P.; Muelhaup, E.; Humbel, R.E.  
Hoppe-Seyler's Z. Physiol. Chem. 253, 451-458, 1972  
A:Title: Amino acid sequence of the two insulins from mouse (Mus musculus).  
A:Reference number: A01592; MUID:72189455; PMID:5063718  
A:Accession: B01592  
A:Molecule type: protein  
A:Residues: 25-54;90-110 <BUE>  
C:Genetics:  
A:Introns: 63/1  
C:Superfamily: insulin  
C:Keywords: hormone; pancreas  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-54/Domain: insulin chain B #status experimental <BCH>  
F:57-87/Domain: connecting peptide #status experimental <MAT>  
F:90-110/Domain: insulin chain A #status experimental <ACH>  
F:31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 85.1%; Score 394; DB 1; Length 110;  
Best Local Similarity 84.9%; Pred. No. 9.2e-36;  
Matches 73; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVYALVLCGERGFFYTPKTRAEADLQVGVGVGGPGAGSLQPLALEG 60  
DB 25 FVNOHLCGSHLVYALVLCGERGFFYTPKTRAEADLQVGVGVGGPGAGSLQPLALEG 84  
QY 61 SLQKRGIVQCCTSCISLYOLENYCN 86  
DB 85 ARQKRGIVQCCTSCISLYOLENYCN 110



```

RESULT 10
A:Residues: 1-110 <COR>
A:Cross-references: GB:J00747; NID:q204956; PIDN:AAA1442.1; PID:q204957
R:Lomedico, P.; Rosenthal, N.; Efstathiadis, A.; Gilbert, W.; Kolodner, R.; Lizard,
Cell 18, 545-558, 1979
A:Title: The structure and evolution of the two nonallelic rat preproinsulin genes.
A:Reference number: A90789; MUID:80045035; PMID:498284
A:Accession: A90789
A:Molecule type: DNA
A:Residues: 1-110 <COM>
A:Cross-references: GB:J00747; NID:q204956; PIDN:AAA1442.1; PID:q204957
R:Steiner, D.F.; Clark, J.L.; Nolan, C.; Rubenstein, A.H.; Margoliash, E.; Aten, B.
Recent Prog. Horm. Res. 25, 207-282, 1969
A:Title: Proinsulin and the biosynthesis of insulin.
A:Reference number: A94231; MUID:70067613; PMID:4111938
A:Accession: A94231
A:Molecule type: protein
A:Residues: 25-54; 90-110 <STE>
R:Tager, H.S.; Steiner, D.F.
J. Biol. Chem. 247, 7936-7940, 1972
A:Title: Primary structures of the proinsulin connecting peptides of the rat and ho
A:Reference number: A92120; MUID:7361499; PMID:440931
A:Accession: B92120
A:Molecule type: protein
A:Residues: 57-87 <IAG>
R:Lomedico, P.T.; Rosenthal, N.; Kolodner, R.; Efstathiadis, A.; Gilbert, W.
Ann. N. Y. Acad. Sci. 443, 425-432, 1980
A:Title: The structure of rat preproinsulin genes.
A:Reference number: 151945; MUID:80240379; PMID:6249167
A:Accession: 151945
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-110 <RES>
A:Cross-references: GB:M25584; NID:q204947; PIDN:AAA1449.1; PID:q204948
C:Genetics:
C:Gene: INS1
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-24/Domain: signal sequence *status predicted <SIG>
F:25-54/Domain: insulin chain B *status experimental <BCR>
F:57-87/Domain: connecting peptide *status experimental <MAT>
F:90-110/Domain: insulin chain A *status experimental <ACH>
F:1-96.43-109.95-100/bisulfide bonds: *status experimental
Query Match 83.2%; Score 385, Db 1; Length 110;
Best Local Similarity 83.7%; Pred. No. 8.8e-45;
Matches 72; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTREFAELQVQLVQVGGVPGGAGSLQPLALEG 60
DB 25 FVQHLCGPHLVEALYLVCGERGFFYTPKSRREVEDPQVQLQLTALAIEV 84

QY 61 SLQKRGIVDQCCTSLCSLYOLENYCN 86
DB 85 ARQKRGIVDQCCTSLCSLYOLENYCN 110

RESULT 13
IPPC
insulin precursor - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 22-Jun-1981 *sequence_revision 22-Jun-1981 *text_change 16-Jul-1999
C:Accession: A01583; A94572; S16492; A60835; B60815
R:Chance, R.E.; Ellis, R.M.; Bromer, W.W.
Science 161, 165-167, 1968
A:Title: Porcine proinsulin: characterization and amino acid sequence.
A:Reference number: A94240; MUID:68286485; PMID:5657063
A:Accession: A01583
A:Molecule type: protein
A:Residues: 1-34, 'Q', 36-84 <CHA>
R:Chance, R.E.
submitted to the Atlas, July 1970
A:Reference number: A94572

```

```

RESULT 10
A:Residues: 1-110 <COR>
A:Cross-references: GB:J00747; NID:q204956; PIDN:AAA1442.1; PID:q204957
R:Lomedico, P.; Rosenthal, N.; Efstathiadis, A.; Gilbert, W.; Kolodner, R.; Lizard,
Cell 18, 545-558, 1979
A:Title: The structure and evolution of the two nonallelic rat preproinsulin genes.
A:Reference number: A90789; MUID:80045035; PMID:498284
A:Accession: A90789
A:Molecule type: DNA
A:Residues: 1-110 <COM>
A:Cross-references: GB:J00747; NID:q204956; PIDN:AAA1442.1; PID:q204957
R:Steiner, D.F.; Clark, J.L.; Nolan, C.; Rubenstein, A.H.; Margoliash, E.; Aten, B.
Recent Prog. Horm. Res. 25, 207-282, 1969
A:Title: Proinsulin and the biosynthesis of insulin.
A:Reference number: A94231; MUID:70067613; PMID:4111938
A:Accession: A94231
A:Molecule type: protein
A:Residues: 25-54; 90-110 <STE>
R:Tager, H.S.; Steiner, D.F.
J. Biol. Chem. 247, 7936-7940, 1972
A:Title: Primary structures of the proinsulin connecting peptides of the rat and immu
A:Reference number: A92120; MUID:7361499; PMID:440931
A:Accession: B92120
A:Molecule type: protein
A:Residues: 57-87 <IAG>
R:Lomedico, P.T.; Rosenthal, N.; Kolodner, R.; Efstathiadis, A.; Gilbert, W.
Ann. N. Y. Acad. Sci. 443, 425-432, 1980
A:Title: The structure of rat preproinsulin genes.
A:Reference number: 151945; MUID:80240379; PMID:6249167
A:Accession: 151945
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-110 <RES>
A:Cross-references: GB:M26328; NID:gl91420; PIDN:AAA37089.1; PID:q305360
C:Superfamily: insulin
Query Match 84.7%; Score 392, Db 2; Length 110;
Best Local Similarity 84.9%; Pred. No. 1.5e-45;
Matches 73; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTREFAELQVQLVQVGGVPGGAGSLQPLALEG 60
DB 25 FVQHLCGPHLVEALYLVCGERGFFYTPKREFAELQVQLVQVGGVPGGAGSLQPLALEG 84

QY 61 SLQKRGIVDQCCTSLCSLYOLENYCN 86
DB 83 PMQKRGIVDQCCTSLCSLYOLENYCN 100

RESULT 11
insulin precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 02-Jul-1996 *sequence_revision 02-Jul-1996 *text_change 16-Jul-1999
C:Accession: I48166
R:Belk, G.I.; Sanchez-Pescador, R.
Diabetes 44, 297-300, 1984
A:Title: Sequence of a cDNA encoding Syrian hamster preproinsulin.
A:Reference number: I48166; MUID:84133036; PMID:6365663
A:Accession: I48166
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-110 <RES>
A:Cross-references: GB:M26328; NID:gl91420; PIDN:AAA37089.1; PID:q305360
C:Superfamily: insulin
Query Match 84.7%; Score 392, Db 2; Length 110;
Best Local Similarity 84.9%; Pred. No. 1.5e-45;
Matches 73; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTREFAELQVQLVQVGGVPGGAGSLQPLALEG 60
DB 25 FVQHLCGPHLVEALYLVCGERGFFYTPKREFAELQVQLVQVGGVPGGAGSLQPLALEG 84

QY 61 SLQKRGIVDQCCTSLCSLYOLENYCN 86
DB 85 ARQKRGIVDQCCTSLCSLYOLENYCN 110

RESULT 12
IPPC
insulin precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 24-Oct-1981 *sequence_revision 24-Oct-1981 *text_change 24-Sep-1999
C:Accession: A90788; A90789; A94231; B92120; 151945; A01589
R:Cordeli, B.; Bell, G.; Tischer, E.; Denk, F.M.; Ullrich, A.; Pictet, R.; Rutter, W.J.
Cell 18, 543-548, 1979
A:Title: Isolation and characterization of a cloned rat insulin gene.
A:Reference number: A90788; MUID:80045034; PMID:498283
A:Accession: A90788
A:Molecule type: DNA

```

```

A:Accession: A94572
A:Molecule type: protein
A:Residues: 1-84 <CH2>
R:Brown, H.; Sanger, F.; Kitai, R.
Biochem. J. 60, 556-565, 1955
A:Title: The structure of pig and sheep insulins.
A:Reference number: A90344
A:Accession: S16492
A:Molecule type: protein
A:Residues: 1-30, 31-51 <BRO>
R:Snell, L.; Damgaard, U.
Horm. Metab. Res. 20, 476-480, 1988
A:Title: Proinsulin heterogeneity in pigs.
A:Reference number: A60835; MUID:89032178; PMID:181865
A:Accession: A60835
A:Molecule type: protein
A:Residues: 33-38, 40-62 <SNE>
A:Note: the authors report the characterization of a connecting peptide variant lacking
A:Molecule type: protein
A:Residues: 33-62 <SN2>
R:Blundell, P.; Dodson, G.; Holstick, P.; Moras, D.
Adv. Protein Chem. 26, 279-402, 1972
A:Title: Insulin: the structure in the crystal and its reflection in chemistry and biology
A:Reference number: A90017
A:Contents: annotation: X-ray crystallography, 1.9 angstroms
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: insulin chain B #status experimental <BCH>
F:30-64-84/Product: insulin #status experimental <MAT>
F:33-63/Domain: connecting peptide #status experimental <CPEP>
F:64-84/Domain: insulin chain A #status experimental <ACH>
F:7-70, 19-83, 69-74/Disulfide bonds: #status experimental

Query Match 82.7%; Score 381; DB 1; Length 84;
Best Local Similarity 86.0%; Pred. No. 10; 34;
Matches 74; Conservative 1; Mismatches 9; Indels 2; Gaps 1;

QY 1 FVNHLCGSHLVGGERGFFYTPK1 REAEDLVGVGVGGPGAGSLOPLALEG 60
|||||
DB 1 FVNHLCGSHLVGGERGFFYTPK1 REAENPVAGAVELGG-GLGGLQALALEG 58
|||||

QY 61 SLQKRGIVEQCCTSGISLYOLENYCN 86
|||||
DB 59 PPQKRGIVEQCCTSGISLYOLENYCN 44
|||||

RESULT 14
IPBO
Insulin precursor - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 24-Apr-1984 #sequence, revision 22-Apr-1995 #text, change 16-Jul-1999
C:Accession: A40909; A92080; A92074; A91185; A90342; A90341; S48184; S48258; A01
R:D'Agostino, J.; Yunes, M.A.; White, J.W.; Hesch, P.K.; Field, J.B.; Frazier, H.L.
Mol. Endocrinol. 1, 327-331, 1987
A:Title: Cloning and nucleotide sequence analysis of complementary deoxyribonucleic acid
A:Reference number: A40909; MUID:88268204; PMID:2456452
A:Accession: A40909
A:Molecule type: mRNA
A:Residues: 1-105 <DAA>
A:Cross-references: GR-M54979; NID-Q163578; P:DN:AAA30722.1; PID:Q163579
A:Experimental source: fetal pancreas
R:Nolan, C.; Marcolliash, E.; Peterson, J.B.; Steiner, D.F.
J. Biol. Chem. 246, 2780-2795, 1971
A:Title: The structure of bovine proinsulin
A:Reference number: A92080; MUID:71166442; PMID:4528892
A:Accession: A92080
A:Molecule type: protein
A:Residues: 25-105 <NOL>
R:Steiner, D.F.; Cho, S.; Oyer, P.E.; Terris, S.; Peterson, J.D.; Rubenstein, A.H.
J. Biol. Chem. 246, 1365-1374, 1971
A:Title: Isolation and characterization of proinsulin C-peptide from bovine pancreas.
A:Reference number: A92074; MUID:7116409; PMID:5545080

```

---

```

A:Accession: A92074
A:Molecule type: protein
A:Residues: 57-82 <STE>
R:Salokangas, A.; Smyth, D.G.; Markussen, J.; Sundby, F.
Eur. J. Biochem. 20, 183-189, 1971
A:Title: Bovine proinsulin: amino acid sequence of the C-peptide isolated from pancre
A:Reference number: A91185; MUID:71257721; PMID:5105368
A:Accession: A91185
A:Molecule type: protein
A:Residues: 57-82 <SAL>
R:Sanger, F.; Thompson, E.O.P.
Biochem. J. 53, 366-374, 1953
A:Title: The amino-acid sequence in the diacyl chain of insulin. 2. The investigatio
A:Reference number: A90342
A:Accession: A90342
A:Molecule type: protein
A:Residues: 85-105 <SAN>
R:Sanger, F.; Tuppy, H.
Biochem. J. 49, 481-490, 1953
A:Title: The amino-acid sequence in the phenylalanyl chain of insulin. 2. The invest
A:Reference number: A90341
A:Accession: A90341
A:Molecule type: protein
A:Residues: 25-54 <SA2>
R:Cheng, R.; Kawakishi, S.
Eur. J. Biochem. 223, 759-764, 1994
A:Title: Site-specific oxidation of histidine residues in glycosylated insulin mediated
A:Reference number: S48184; MUID:94333378; PMID:8055951
A:Accession: S48184
A:Molecule type: protein
A:Residues: 85-105 <CHE>
A:Accession: S48185
A:Status: preliminary
A:Molecule type: protein
A:Residues: 25-30, 32-42, 44-54 <CH2>
R:Ryle, A.P.; Sanger, F.; Smith, L.F.; Kitai, R.
Biochem. J. 60, 541-556, 1955
A:Title: The disulphide bonds of insulin.
A:Reference number: A90343
A:Contents: annotation: amides; disulfides
R:Wenzel, T.; Eckerskorn, C.; Lottspeich, F.; Haunmeister, W.
FEBS Lett. 349, 205-209, 1994
A:Title: Existence of a molecular ruler in proinsulins suggested by analysis of deq
A:Reference number: S46258; MUID:94326921; PMID:8050567
A:Accession: S46258
A:Status: preliminary
A:Molecule type: protein
A:Residues: 25-54 <WEN>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-54/Domain: insulin chain B #status experimental <BCH>
F:25-54, 85-105/Product: insulin #status experimental <MAT>
F:57-82/Domain: connecting peptide #status experimental <CPEP>
F:85-105/Domain: insulin chain A #status experimental <ACH>
F:31-91, 43-104, 90-95/Disulfide bonds: #status experimental

Query Match 79.2%; Score 366.5; DB 1; Length 105;
Best Local Similarity 80.2%; Pred. No. 6; 7e-31;
Matches 69; Conservative 2; Mismatches 10; Indels 5; Gaps 1;

QY 1 FVNHLCGSHLVGGERGFFYTPK1 REAEDLVGVGVGGPGAGSLOPLALEG 60
|||||
DB 25 FVNHLCGSHLVGGERGFFYTPK1 REAENPVAGAVELGG-GLGGLQALALEG 58
|||||

QY 61 SLQKRGIVEQCCTSGISLYOLENYCN 86
|||||
DB 80 PPQKRGIVEQCCTSGISLYOLENYCN 105
|||||

RESULT 15
INMS1
Insulin 1 precursor - mouse

```

C:Species: Mus musculus (house mouse)  
 C:Date: 24-Apr-1984 #sequence\_revision 14-Jul-1994 #text\_change 18-Jun-1999  
 C:Accession: B26342; A48172; A01592; B61012  
 R:Wentworth, B.M.; Schaefer, I.M.; Villa-Komaroff, L.; Chirgwin, J.M.  
 J. Mol. Evol. 23, 305-312, 1986  
 A:Title: Characterization of the two nonallelic genes encoding mouse preproinsulin.  
 A:Reference number: A92965; MUID:87169768; PMID:3104603  
 A:Accession: B26342  
 A:Molecule type: DNA  
 A:Residues: 1108 <MEN>  
 A:Cross-references: GB:X04725; NID:q52712; PIDN:CAA28434.1; PID:q52714  
 R:Sawa, T.; Onakura, S.; Morioka, H.; Yano, S.  
 J. Mol. Endocrinol. 5, 61-67, 1990  
 A:Title: Molecular cloning and DNA sequence analysis of preproinsulin genes in the NON m  
 A:Reference number: A48172; MUID:90372989; PMID:2397023  
 A:Accession: A48172  
 A>Status: not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1108 <SAW>  
 R:Acenzio, A.F.; Balthazar, R.; Kunz, P.; Moellhaupt, E.; Hamborg, F.E.  
 Hoppe Seyler's Z. Physiol. Chem. 363, 451-456, 1972  
 A:Title: Amino acid sequence of the two insulin gene copies (Mus musculus)  
 A:Reference number: A01592; MUID:7276455; PMID:5004718  
 A:Accession: A01592  
 A:Molecule type: protein  
 A:Residues: 25 54:88-108 <HUE>  
 R:Lindø, S.; Nielsen, J.H.; Hansen, B.; Molinder, B.S.  
 J. Chromatogr. 462, 243-254, 1989  
 A:Title: Reversed-phase high-performance liquid chromatographic analyses of insulin bios  
 A:Reference number: A61012; MUID:89292678; PMID:2661585  
 A:Accession: B61012  
 A:Molecule type: protein  
 A:Residues: 57-85 <LIN>  
 C:Superfamily: insulin  
 C:Keywords: hormone; pancreas  
 F:1-24/domain: signal sequence #status predicted <SIG>  
 F:25-54/domain: insulin chain B #status experimental <BCB>  
 F:57-85/domain: connecting peptide #status experimental <CAP>  
 F:88-108/domain: insulin chain A #status experimental <ACH>  
 F:31-94,44-107,93-98/Disulfide bonds: #status predicted

Query Match 79.08; Score 366; DB 1; Length 108;

Best Local Similarity 81.4%; Pred. No. 16-32; Indels 2; Gaps 1;

Matches 70; Conservative 4; Mismatches 10;

QY 1 FVNHLCGSHLVEALYVGCGEGFFYTPKTRREAEDLQVGQVGGPGAGSLQPLAIEG 60

DB 25 FVKGHLCGPHLVEALYVGCGEGFFYTPKSRREVEDNVEQELGCGSP--GDLQTLAIEV 82

QY 61 SLQKRRIIVQCCCTSCSLYQLFNQCN 86

DB 83 ANQKRRIIVQCCCTSCSLYQLFNQCN 108

Search completed: September 15, 2003, 12:04:19  
 Job time : 34.1792 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:54:00 : Search time 8.01434 Seconds  
(without alignments)  
504.633 Million cell updates/sec

Title: US-09-423-100-4

Perfect score: 463

Sequence: 1 FVNQHLGSHLVEALYLGV.....IVEQCTSIQSLYLENYCN R6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127864

Minimum hit seq length: 0

Maximum hit seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY:

Result No.	Score	Query Match	Length	DB ID	Description
1	453	100.0	110	1	INS_HUMAN
2	454	100.0	110	1	INS_PANTR
3	456	98.5	110	1	INS_CERAE
4	456	98.5	110	1	INS_MACFA
5	424	91.6	110	1	INS_RABIT
6	417	90.1	110	1	INS_CANFA
7	413	89.2	110	1	INS_SHEPR
8	394	85.1	86	1	INS_HORSE
9	394	85.1	110	1	INS2_MOUSE
10	394	85.1	110	1	INS2_RAT
11	392	84.7	108	1	INS_AOTIR
12	392	84.7	110	1	INS_CRILA
13	385	83.2	110	1	INS1_RAT
14	383	82.7	108	1	INS_PLIG
15	377	81.4	110	1	INS_PSOR
16	366.5	79.2	105	1	INS_BOVIN
17	366	79.0	108	1	INS1_MOUSE
18	362.5	78.3	105	1	INS_SHEEP
19	334.5	72.2	108	1	INS_RDOSP
20	314	67.8	110	1	INS_CAVPO
21	277.5	59.9	109	1	INS_OTDE
22	276.5	59.7	103	1	INS_SELF
23	265.5	57.3	106	1	INS2_XENLA
24	265.5	57.3	107	1	INS_CHICK
25	262.5	56.7	106	1	INS1_XENLA
26	256.5	55.4	51	1	INS_BALPH
27	256.5	55.4	51	1	INS_ELENA
28	256.5	55.4	81	1	INS_ANAPL
29	251.5	54.3	51	1	INS_ACOCA
30	246.5	53.2	51	1	INS_BALBO
31	246.5	53.2	51	1	INS_CAMDR
32	246.5	53.2	51	1	INS_CAPHI
33	245.5	53.0	51	1	INS_FELCA

34 239.5 51.7 51 1 INS\_DIDMA  
35 234.5 50.6 51 1 INS\_CHIBR  
36 231.5 50.0 51 1 INS\_ANGAN  
37 228 49.2 115 1 INS\_VERMO  
38 227.5 49.1 51 1 INS\_HYSCR  
39 227.5 49.1 51 1 INS\_TRASC  
40 225 48.6 52 1 INS\_ACIGU  
41 221.5 47.8 116 1 INS\_LOPPI  
42 218.5 47.2 51 1 INS\_ORNAN  
43 218.5 47.2 105 1 INS\_ONCKE  
44 216.5 46.8 89 1 INS\_CALMI  
45 215 46.4 113 1 INS\_ORENI

# ALIGNMENTS

RESULT 1  
INS\_HUMAN STANDARD; ERT: 110 AA.  
AC P01308;  
DE 21-001-1986 (Ref: 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, 1000, 1001, 1002, 1003, 1004, 1005, 1006, 1007, 1008, 1009, 1010, 1011, 1012, 1013, 1014, 1015, 1016, 1017, 1018, 1019, 1020, 1021, 1022, 1023, 1024, 1025, 1026, 1027, 1028, 1029, 1030, 1031, 1032, 1033, 1034, 1035, 1036, 1037, 1038, 1039, 1040, 1041, 1042, 1043, 1044, 1045, 1046, 1047, 1048, 1049, 1050, 1051, 1052, 1053, 1054, 1055, 1056, 1057, 1058, 1059, 1060, 1061, 1062, 1063, 1064, 1065, 1066, 1067, 1068, 1069, 1070, 1071, 1072, 1073, 1074, 1075, 1076, 1077, 1078, 1079, 1080, 1081, 1082, 1083, 1084, 1085, 1086, 1087, 1088, 1089, 1090, 1091, 1092, 1093, 1094, 1095, 1096, 1097, 1098, 1099, 1100, 1101, 1102, 1103, 1104, 1105, 1106, 1107, 1108, 1109, 1110, 1111, 1112, 1113, 1114, 1115, 1116, 1117, 1118, 1119, 1120, 1121, 1122, 1123, 1124, 1125, 1126, 1127, 1128, 1129, 1130, 1131, 1132, 1133, 1134, 1135, 1136, 1137, 1138, 1139, 1140, 1141, 1142, 1143, 1144, 1145, 1146, 1147, 1148, 1149, 1150, 1151, 1152, 1153, 1154, 1155, 1156, 1157, 1158, 1159, 1160, 1161, 1162, 1163, 1164, 1165, 1166, 1167, 1168, 1169, 1170, 1171, 1172, 1173, 1174, 1175, 1176, 1177, 1178, 1179, 1180, 1181, 1182, 1183, 1184, 1185, 1186, 1187, 1188, 1189, 1190, 1191, 1192, 1193, 1194, 1195, 1196, 1197, 1198, 1199, 1200, 1201, 1202, 1203, 1204, 1205, 1206, 1207, 1208, 1209, 1210, 1211, 1212, 1213, 1214, 1215, 1216, 1217, 1218, 1219, 1220, 1221, 1222, 1223, 1224, 1225, 1226, 1227, 1228, 1229, 1230, 1231, 1232, 1233, 1234, 1235, 1236, 1237, 1238, 1239, 1240, 1241, 1242, 1243, 1244, 1245, 1246, 1247, 1248, 1249, 1250, 1251, 1252, 1253, 1254, 1255, 1256, 1257, 1258, 1259, 1260, 1261, 1262, 1263, 1264, 1265, 1266, 1267, 1268, 1269, 1270, 1271, 1272, 1273, 1274, 1275, 1276, 1277, 1278, 1279, 1280, 1281, 1282, 1283, 1284, 1285, 1286, 1287, 1288, 1289, 1290, 1291, 1292, 1293, 1294, 1295, 1296, 1297, 1298, 1299, 1300, 1301, 1302, 1303, 1304, 1305, 1306, 1307, 1308, 1309, 1310, 1311, 1312, 1313, 1314, 1315, 1316, 1317, 1318, 1319, 1320, 1321, 1322, 1323, 1324, 1325, 1326, 1327, 1328, 1329, 1330, 1331, 1332, 1333, 1334, 1335, 1336, 1337, 1338, 1339, 1340, 1341, 1342, 1343, 1344, 1345, 1346, 1347, 1348, 1349, 1350, 1351, 1352, 1353, 1354, 1355, 1356, 1357, 1358, 1359, 1360, 1361, 1362, 1363, 1364, 1365, 1366, 1367, 1368, 1369, 1370, 1371, 1372, 1373, 1374, 1375, 1376, 1377, 1378, 1379, 1380, 1381, 1382, 1383, 1384, 1385, 1386, 1387, 1388, 1389, 1390, 1391, 1392, 1393, 1394, 1395, 1396, 1397, 1398, 1399, 1400, 1401, 1402, 1403, 1404, 1405, 1406, 1407, 1408, 1409, 1410, 1411, 1412, 1413, 1414, 1415, 1416, 1417, 1418, 1419, 1420, 1421, 1422, 1423, 1424, 1425, 1426, 1427, 1428, 1429, 1430, 1431, 1432, 1433, 1434, 1435, 1436, 1437, 1438, 1439, 1440, 1441, 1442, 1443, 1444, 1445, 1446, 1447, 1448, 1449, 1450, 1451, 1452, 1453, 1454, 1455, 1456, 1457, 1458, 1459, 1460, 1461, 1462, 1463, 1464, 1465, 1466, 1467, 1468, 1469, 1470, 1471, 1472, 1473, 1474, 1475, 1476, 1477, 1478, 1479, 1480, 1481, 1482, 1483, 1484, 1485, 1486, 1487, 1488, 1489, 1490, 1491, 1492, 1493, 1494, 1495, 1496, 1497, 1498, 1499, 1500, 1501, 1502, 1503, 1504, 1505, 1506, 1507, 1508, 1509, 1510, 1511, 1512, 1513, 1514, 1515, 1516, 1517, 1518, 1519, 1520, 1521, 1522, 1523, 1524, 1525, 1526, 1527, 1528, 1529, 1530, 1531, 1532, 1533, 1534, 1535, 1536, 1537, 1538, 1539, 1540, 1541, 1542, 1543, 1544, 1545, 1546, 1547, 1548, 1549, 1550, 1551, 1552, 1553, 1554, 1555, 1556, 1557, 1558, 1559, 1560, 1561, 1562, 1563, 1564, 1565, 1566, 1567, 1568, 1569, 1570, 1571, 1572, 1573, 1574, 1575, 1576, 1577, 1578, 1579, 1580, 1581, 1582, 1583, 1584, 1585, 1586, 1587, 1588, 1589, 1590, 1591, 1592, 1593, 1594, 1595, 1596, 1597, 1598, 1599, 1600, 1601, 1602, 1603, 1604, 1605, 1606, 1607, 1608, 1609, 1610, 1611, 1612, 1613, 1614, 1615, 1616, 1617, 1618, 1619, 1620, 1621, 1622, 1623, 1624, 1625, 1626, 1627, 1628, 1629, 1630, 1631, 1632, 1633, 1634, 1635, 1636, 1637, 1638, 1639, 1640, 1641, 1642, 1643, 1644, 1645, 1646, 1647, 1648, 1649, 1650, 1651, 1652, 1653, 1654, 1655, 1656, 1657, 1658, 1659, 1660, 1661, 1662, 1663, 1664, 1665, 1666, 1667, 1668, 1669, 1670, 1671, 1672, 1673, 1674, 1675, 1676, 1677, 1678, 1679, 1680, 1681, 1682, 1683, 1684, 1685, 1686, 1687, 1688, 1689, 1690, 1691, 1692, 1693, 1694, 1695, 1696, 1697, 1698, 1699, 1700, 1701, 1702, 1703, 1704, 1705, 1706, 1707, 1708, 1709, 1710, 1711, 1712, 1713,

- RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullighy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Guarnarone P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay E.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren F.J., Lu X., Gibbs R.A.,  
RA Farley J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Blakesley R.W., Young A.C., Shevchenko V., Bouffard G.G.,  
RA Rodriguez A.C., Grimwood J., Green E.D., Dickson M.C.,  
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Morra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [7]  
RN SEQUENCE OF 1-59 FROM N.A.  
RN TISSUE-BLOOD;  
RN Fajardy I.L., Weill J.J., Stuckens C.C., Danze P.M.P.;  
RT "Description of a novel RFLP diallelic polymorphism (-127 BspI C/G;  
RT within the 5' region of insulin gene.";  
RL Submitted (JUL-1998) to the EMBL/GenBank/  
RN [8]  
RN SEQUENCE OF 25-54 AND 90-111;  
RN Nicol D.S.B.W., Smith L.F.;  
RT "Amino-acid sequence of human insulin.";  
RL Nature 187:483-485(1960).  
RN [9]  
RN SEQUENCE OF 57-87.  
RN OVER P.E., Cho S., Peterson J.D., Steh D.F.;  
RA "Studies on human proinsulin. Isolation of amino acid sequence of  
RT the human pancreatic C-peptide.";  
RL J. Biol. Chem. 246:1375-1386(1971).  
RN [10]  
RN SEQUENCE OF 57-87.  
RN MEDLINE-71257722; PubMed-5560404;  
RA Ko A., Smyth D.G., Markussen J., Sundby I.;  
RT "The amino acid sequence of the C-peptide of human proinsulin.";  
RL Eur. J. Biochem. 20:190-199(1971).  
RN [11]  
RN SYNTHESIS;  
RN MEDLINE-7507277; PubMed-4433293;  
RA Sieber P., Kamber B., Hartmann A., Joehli A., Riniker B., Rittel W.;  
RT "Total synthesis of human insulin under directed formation of the  
RT disulfide bonds.";  
RL Helv. Chim. Acta 57:2617-2621(1974).  
RN [12]  
RN SYNTHESIS OF 57-87.  
RN MEDLINE-75040007; PubMed-4803504;  
RA Naithani V.K.;  
RT "Studies on polypeptides. IV. The synthesis of C-peptide of human  
RT proinsulin.";  
RL Hoppe-Seyler's Z. Physiol. Chem. 354:659-672(1973).  
RN [13]  
RN SYNTHESIS OF 65-69 AND 70-73  
RN MEDLINE-73161263; PubMed-4698555;  
RA Geiger P., Volk A.;  
RT "Synthesis of peptides with the properties of human proinsulin C  
RT peptides (hC peptide). 3. Synthesis of the sequences 14-17 and 9-13  
RT of human proinsulin C-peptides.";  
RL Chem. Ber. 106:199-205(1973).  
RN [14]  
RN SYNTHESIS OF 84-87.  
RN MEDLINE-73161261; PubMed-4698551;  
RA Geiger P., Jaeger G., Keonig W., Treutlein G.;  
RT "Synthesis of peptides with the properties of human proinsulin C  
RT peptides (hC peptide). 1. Scheme for the synthesis and preparation of  
RT the sequence 28-31 of human proinsulin C-peptide.";  
RL Chem. Ber. 106:188-192(1973).  
RN [15]  
RN VARIANT LOS ANGELES SER-48.  
RN MEDLINE-84016053; PubMed-6312455;  
RA Haneda M., Chan S.-J., Kwok S.C.M., Rubenstein A.H., Steiner D.F.;  
RT "Studies on mutant human insulin genes: identification and sequence  
RT analysis of a gene encoding [SerB24]insulin.";  
RL Proc. Natl. Acad. Sci. U.S.A. 80:6366-6370(1983).  
RN [16]  
RN VARIANTS LOS ANGELES SER-48 AND CHICAGO LEU-49.  
RN MEDLINE-84170233; PubMed-6424111;  
RA Shoelson S., Fickova M., Haneda M., Musso G., Kaiser E.T.,  
RA Rubenstein A.H., Tager H.;  
RT "Identification of a mutant human insulin predicted to contain a  
RT serine-for-phenylalanine substitution.";  
RL Proc. Natl. Acad. Sci. U.S.A. 80:7390-7394(1983).  
RN [17]  
RN VARIANT PROVIDENCE ASP 34.  
RN MEDLINE-87175640; PubMed-3470784;  
RA Chan S.-J., Seino S., Gruppuso P.A., Schwartz R., Steiner D.F.;  
RT "A mutation in the B chain coding region is associated with impaired  
RT proinsulin conversion in a family with hyperproinsulinemia.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:2194-2197(1987).  
RN [18]  
RN VARIANT WAKAYAMA LEU 92.  
RN MEDLINE-87058122; PubMed-3537011;  
RA Sakurai H., Iwamoto Y., Sakamoto Y., Kuzuya T., Hirata H.;  
RT "Structurally abnormal insulin in a diabetic patient. Characterization  
RT of the mutant insulin A3 (Val-->Leu) isolated from the pancreas.";  
RL J. Clin. Invest. 78:1466-1472(1986).  
RN [19]  
RN VARIANT HIS-89.  
RN MEDLINE-90317021; PubMed-2196279;  
RA Barbetti F., Raben N., Kadowaki T., Cama A., Accili D., Gabbay K.H.,  
RA Merenich J.A., Taylor S.L., Roth J.;  
RT "Two unrelated patients with familial hyperproinsulinemia due to a  
RT mutation substituting histidine for arginine at position 65 in the  
RT proinsulin molecule: identification of the mutation by direct  
RT sequencing of genomic deoxyribonucleic acid amplified by polymerase  
RT chain reaction.";  
RL J. Clin. Endocrinol. Metab. 71:164-169(1990).  
RN [20]  
RN VARIANT HIS-89.  
RN MEDLINE-85261996; PubMed-4019786;  
RA Shibasaki Y., Kawakami T., Kanazawa Y., Akanuma Y., Takaku F.;  
RT "Posttranslational cleavage of proinsulin is blocked by a point  
RT mutation in familial hyperproinsulinemia.";  
RL J. Clin. Invest. 76:378-380(1985).  
RN [21]  
RN VARIANT KYOTO LEU-89.  
RN MEDLINE-92291307; PubMed-1601997;  
RA Yano H., Kitano N., Morimoto M., Polonsky K.S., Imura H., Seino Y.;  
RT "A novel point mutation in the human insulin gene giving rise to  
RT hyperproinsulinemia (proinsulin Kyoto).";  
RL J. Clin. Invest. 89:1902-1907(1992).  
RN [22]  
RN STRUCTURE BY NMR.  
RN MEDLINE-91104966; PubMed-2271664;  
RA Hua Q.-X., Weiss M.A.;  
RT "Toward the solution structure of human insulin: sequential 2D NMR  
RT assignment of a des-pentapeptide analogue and comparison with crystal  
RT structure.";  
RL Biochemistry 29:10545-10555(1990).  
RN [23]  
RN STRUCTURE BY NMR.  
RN MEDLINE-91242467; PubMed-2036420;  
RA Hua Q.-X., Weiss M.A.;  
RT "Comparative 2D NMR studies of human insulin and des-pentapeptide  
RT insulin: sequential resonance assignment and implications for protein  
RT dynamics and receptor recognition.";  
RL Biochemistry 30:5505-5515(1991).  
RN [24]  
RN STRUCTURE BY NMR.  
RN MEDLINE-91265527; PubMed-1646635;  
RA Hua Q.-X., Weiss M.A.;  
RT "Two-dimensional NMR studies of Des-(B26-B30)-insulin: sequence-  
RT specific resonance assignments and effects of solvent composition.";  
RL Biochim. Biophys. Acta 1078:101-110(1991).

Query Match 100.0%; Score 463; DB 1; Length 110;  
 Best local Similarity 100.0%; Pred. No. 9.2e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 60  
 DB 25 FVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 84  
 QY 61 SLQKRGIVEQCCTSTCSLYQLENYCN 86  
 DB 85 SLQKRGIVEQCCTSTCSLYQLENYCN 110

## RESULT 2

INS\_PANTR ID INS\_PANTR STANDARD; PRT: 110 AA;  
 AC P30410;  
 DT 01-APR-1993 (rel. 25, Created);  
 DT 01-APR-1993 (rel. 25, Last sequence update);  
 DT 15-SEP-2003 (rel. 42, Last annotation update);  
 DE Insulin precursor;  
 GN INS;  
 OS Pan troglodytes (Hominidae);  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan;  
 OX NCBI\_TaxID=9598;  
 RN [1];  
 RP SEQUENCE FROM N.A.

RX MEDLINE:92219553; PubMed:1560757;  
 RA Seino S., Bell G.I., Li W.;  
 RT "Sequences of primate insulin genes support the hypothesis of a  
 RT slower rate of molecular evolution in humans and apes than in  
 RT monkeys";  
 RL Mol. Biol. Evol. 9:193-203(1992);  
 CC SEQUENCE OF 57-87.

CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.

CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- SIMILARITY: BELONGS TO THE INSULIN CT/RELAXIN FAMILY.

CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X61089; CAA43403.1;

DR PIR: A42179; A42179.

DR PIR: JEFFE; 25 MAR-00.

DR Pfam: PF00049; Insulin; 1.

DR PRINTS: PR00277; INSULINB.

DR SMART: SM00078; IIGF; 1.

DR PROSITE: PS00262; INSULIN; 1.

DR Insulin family; Hormone; Glucose metabolism; Signal; 3D-structure.

FT SIGNAL 1 24

FT CHAIN 25 54 INSULIN B CHAIN.

FT PROPEP 57 87 C PEPTIDE.

FT CHAIN 90 110 INSULIN A CHAIN.

FT DISULFID 31 96 INTERCHAIN.

FT DISULFID 43 109 INTERCHAIN.

FT DISULFID 95 100

FT SEQUENCE 110 AA; 12025 MW; 41E880F79837CEFS CRC64;

Query Match 100.0%; Score 463; DB 1; Length 110;

Best local Similarity 100.0%; Pred. No. 9.2e-43;

Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 60  
 DB 25 FVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 84  
 QY 61 SLQKRGIVEQCCTSTCSLYQLENYCN 86  
 DB 85 SLQKRGIVEQCCTSTCSLYQLENYCN 110

## RESULT 3

INS\_CERAE ID INS\_CERAE STANDARD; PRT: 110 AA;  
 AC P30407; P01309;  
 DT 01-APR-1993 (rel. 25, Created);  
 DT 01-APR-1993 (rel. 25, Last sequence update);  
 DT 01-FEB-1996 (rel. 31, Last annotation update);  
 DE Insulin precursor;  
 GN INS.

OS Cercopithecus aethiops (Green monkey) (Grivet);  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 CC Cercopithecinae; Peropithecus;  
 OX NCBI\_TaxID=9554;  
 RN [1];  
 RP SEQUENCE FROM N.A.

RX MEDLINE:92219553; PubMed:1560757;

RA Seino S., Bell G.I., Li W.

RT "Sequences of primate insulin genes support the hypothesis of a  
 RT slower rate of molecular evolution in humans and apes than in  
 RT monkeys";  
 RL Mol. Biol. Evol. 9:193-203(1992);  
 CC SEQUENCE OF 57-87.

RX MEDLINE:72258016; PubMed:4626369;

RA Peterson J.D., Nehrlich S., Oyer P.E., Steiner D.F.;

RT "Determination of the amino acid sequence of the monkey, sheep, and  
 RT dog proinsulin C-peptides by a semi-micro Edman degradation  
 RT procedure";  
 RL J. Biol. Chem. 247:4846-4871(1972);

CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.

CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X61092; CAA43405.1;

DR PIR: H42179; H42179.

DR HSSP: P01308; A10.

DR InterPro: IPR004825; Ins/IGF/relax.

DR Pfam: PF00049; Insulin; 1.

DR SMART: SM00078; IIGF; 1.

DR PROSITE: PS00262; INSULIN; 1.

DR Insulin family; Hormone; Glucose metabolism; Signal.

FT SIGNAL 1 24

FT CHAIN 25 54 INSULIN B CHAIN.

FT PROPEP 57 87 C PEPTIDE.

FT CHAIN 90 110 INSULIN A CHAIN.

FT DISULFID 31 96 INTERCHAIN.

FT DISULFID 43 109 INTERCHAIN.

FT DISULFID 95 100

FT SEQUENCE 110 AA; 12019 MW; 95A1F54BE7B247F9 CRC64;

Query Match 98.5%; Score 456; DB 1; Length 110;  
 Best Local Similarity 98.8%; Pred. No. 5.2e-42;  
 Matches 85; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVYALVCGGFFYTPKTRAEADLVQGVGLGGPGAGSLCPALLEG 60  
 DB 25 FVNHLCGSHLVYALVCGGFFYTPKTRAEADLVQGVGLGGPGAGSLCPALLEG 84  
 OY 61 SLQKRGIVQCCTSCSLYQLENYCN 86  
 DB 85 SLQKRGIVQCCTSCSLYQLENYCN 110

RESULT 4

INS\_MACFA STANDARD: PRT: 110 AA.

ID INS\_MACFA  
 AC P01304;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DE 01-FEB-1996 (Rel. 33, Last annotation update)  
 GN INS.  
 OS Erythrocytes (Circulating erythrocytes) (Circulating erythrocytes)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Insuperphala; Leporiformes; Erythrocytes;  
 OX NCBI\_TaxID=9543;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE 81080474; PubMed-6184262;  
 RA Wetekamp W., Gronenberg J., Leineweber M., Wenzenmayer F.,  
 RA Wisnack E.L.;  
 RT \*The nucleotide sequence of cDNA coding for preproinsulin from the  
 RT primate Macaca fascicularis;  
 RL Gene 19:179-183(1982).  
 CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/annouce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: J00336; AAA16849.1;  
 PIR: J00178; J00178.  
 HSSP: P01308; 1A10.  
 InterPro: IPR004825; Ins/IGF/relax.  
 Pfam: PF00049; Insulin; 1.  
 PRINTS: PR00277; INSULIN.  
 SMART: SM00078; IIGF; 1.  
 PROSITE: PS00262; INSULIN; 1.  
 Insulin family; Hormone; Glucose metabolism; Signal.  
 FT CHAIN 1 24  
 FT CHAIN 25 54  
 FT PROPEP 57 87  
 FT CHAIN 90 110

Query Match 98.5%; Score 456; DB 1; Length 110;  
 Best Local Similarity 98.8%; Pred. No. 5.2e-42;  
 Matches 85; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVYALVCGGFFYTPKTRAEADLVQGVGLGGPGAGSLCPALLEG 60  
 DB 25 FVNHLCGSHLVYALVCGGFFYTPKTRAEADLVQGVGLGGPGAGSLCPALLEG 84  
 OY 61 SLQKRGIVQCCTSCSLYQLENYCN 86  
 DB 85 SLQKRGIVQCCTSCSLYQLENYCN 110

RESULT 5

INS\_RABIT STANDARD: PRT: 110 AA.

ID INS\_RABIT  
 AC P01311;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DE 01-OCT-1996 (Rel. 34, Last annotation update)  
 GN INS.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Insuperphala; Leporiformes; Erythrocytes;  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN: New Zealand white; Tissue: Pancreas;  
 RX MEDLINE 94179230; PubMed-8142571;  
 RA Devaskar S.U., Giddings S.J., Rajakumar P.A., Carnaghi L.R.,  
 RA Meon R.K., Zahm D.S.;  
 RT \*Insulin gene expression and insulin synthesis in mammalian neuronal  
 RT cells;  
 RL J. Biol. Chem. 259:8445-8454(1994).  
 RN [2]  
 RP SEQUENCE OF 25-54 AND 90-110.  
 RX MEDLINE 66160119; PubMed-5949593;  
 RA Smith L.F.;  
 RT \*Species variation in the amino acid sequence of insulin.\*;  
 RL Am. J. Med. 40:662-666(1966).  
 RN [3]  
 RP SEQUENCE OF 56-110 FROM N.A.  
 RA Giddings S.J., Carnaghi L.R., Devaskar S.U.;  
 RL Submitted (Apr-1991) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/annouce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: U03610; AAA19044.1;  
 EMBL: M61154; AAA17540.1;  
 PIR: A53438; INRB.  
 HSSP: P01308; IITYM.  
 InterPro: IPR004825; Ins/IGF/relax.  
 Pfam: PF00049; Insulin; 1.  
 SMART: SM00078; IIGF; 1.  
 PROSITE: PS00262; INSULIN; 1.  
 Insulin family; Hormone; Glucose metabolism; Signal.  
 FT CHAIN 1 24  
 FT CHAIN 25 54  
 FT PROPEP 57 87  
 FT CHAIN 90 110

```

FT DISULFID 31 96 INTERCHAIN.
FT DISULFID 43 109 INTERCHAIN.
FT DISULFID 95 100
FT CONFLICT 83 83 E -> Y (IN REF. 3).
SQ SEQUENCE 110 AA: 82D2975B85D77FA8 CRC64:

Query Match 91.6% Score 424: DB 1: Length 110:
Best Local Similarity 90.7% Pred. No. 1.4e-38:
Matches 78: Conservative 3: Mismatches 5: Indels 0: Gaps 0:

QY 1 FVQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLGGPGAGSLQPLEAG 60
   |||||
DQ 25 FVQHLCGSHLVEALYLVCGERGFFYTPKSRREVEELQVQAELOGGPGAGSLQSALE 84
   |||||

QY 61 SLOKRGIVEQCCTSIQSLYLENYCN 86
   |||||
DQ 85 ALOKRGIVEQCCTSIQSLYLENYCN 110

RESULT 6
INS_CANFA
ID INS_CANFA STANDARD EFT: 110 AA:
AC P01321
DT 21-JUL-1986 (Rel. 31, Created)
DT 21-JUL-1986 (Rel. 31, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin precursor.
GN INS.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN 1
RP SEQUENCE FROM N.A.
RX MEDLINE: 83109071; PubMed: 6296142;
RA Kwok S.C.M., Chan S.J., Steiner D.F.;
RT Cloning and nucleotide sequence analysis of the dog insulin gene.
RT Coded amino acid sequence of canine preproinsulin predicts an
RT additional C-peptide fragment.
RL J. Biol. Chem. 258:2357-2363(1983).
RN 2
RP SEQUENCE OF 25-54 AND 90-110.
RX MEDLINE: 66160119; PubMed: 5949594;
RA Smith L.F.;
RT Species variation in the amino acid sequence of insulin.*;
RL Ann. J. Med. 40:662-666(1966).
CC 1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC 2- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC 3- SUBCELLULAR LOCATION: Secreted.
CC 4- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: V00179; CAA23475.1;
DR DR
DR FIC: A92414; IPDG.
DR HSSP: P01317; IAPH.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
DR Insulin family: Hormone; Glucose metabolism; Signal.
DR SIGNAL: 1 24 INSULIN B CHAIN.
FT FT
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 87 C PEPTIDE.
FT CHAIN 90 109 INTERCHAIN.
FT DISULFID 95 100
FT CONFLICT 83 83 E -> Y (IN REF. 3).
SQ SEQUENCE 110 AA: 82D2975B85D77FA8 CRC64:

```

```

FT PROPEP 57 87 C PEPTIDE.
FT CHAIN 90 110 INSULIN A CHAIN.
FT DISULFID 31 96 INTERCHAIN.
FT DISULFID 43 109 INTERCHAIN.
FT DISULFID 95 100
SQ SEQUENCE 110 AA: A574791864A4FB98 CRC64:

Query Match 90.1% Score 417: DB 1: Length 110:
Best Local Similarity 89.5% Pred. No. 7.7e-38:
Matches 77: Conservative 1: Mismatches 8: Indels 0: Gaps 0:

QY 1 FVQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLGGPGAGSLQPLEAG 60
   |||||
DQ 25 FVQHLCGSHLVEALYLVCGERGFFYTPKARREVEELQVQAELOGGPGAGSLQPLEAG 84
   |||||

QY 61 SLOKRGIVEQCCTSIQSLYLENYCN 86
   |||||
DQ 85 ALOKRGIVEQCCTSIQSLYLENYCN 110

RESULT 7
INS_SIBCH
ID INS_SIBCH STANDARD EFT: 110 AA:
AC Q1X13
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Insulin precursor
GN INS.
OS Spermophilus tridecemlineatus (Thirteen-lined ground squirrel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
OC Spermophilus.
OX NCBI_TaxID=43179;
RN 1
RP SEQUENCE FROM N.A.
RX Tredrea M.M., Buck M.J., Guhaniyoni J., Squire T.L., Andrews M.T.;
RT Regulation of PDK4 expression in a hibernating mammal.*;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC 1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC 2- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC 3- SUBCELLULAR LOCATION: Secreted.
CC 4- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: AY038604; AAK72558.1;
DR DR
DR HSSP: P01308; ILNP.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00277; INSULINB.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
DR Insulin family: Hormone; Glucose metabolism; Signal.
DR SIGNAL: 1 24 BY SIMILARITY.
FT FT
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 87 C PEPTIDE.
FT CHAIN 90 110 INSULIN A CHAIN.
FT DISULFID 31 96 INTERCHAIN (BY SIMILARITY).
FT DISULFID 43 109 INTERCHAIN (BY SIMILARITY).
FT DISULFID 95 100 BY SIMILARITY.
SQ SEQUENCE 110 AA: 12004 MW: 4511768D6622BEE5 CRC64:

```



Query Match 89.2% Score 413; DB 1; Length 110;  
 Best Local Similarity 89.5%; Pred. No. 20-37;  
 Matches 77; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLYEALVLCGERGFTYTPKTRAEADLOVGGVGLGGPGAGSLQPIALEG 60  
 DB 25 FVNHLCGSHLYEALVLCGERGFTYTPKTRAEADLOVGGVGLGGPGAGSLQPIALEM 84

QY 61 SLOKRGIVEOCCTSCISLYOLENYCN 86  
 DB 85 ALOKRGIVEOCCTSCISLYOLENYCN 110

RESULT 8  
 INS\_HORSE STANDARD; PKT; 86 AA.  
 AC P01310;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1994 (Rel. 38, Last annotation update)  
 DE Insulin precursor  
 GN INS  
 OS Equus caballus (horse)  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Equi; Equus  
 OX NCBI\_TaxID=9796;  
 RN [1]  
 RP SEQUENCE OF 1-30 AND 56-86.  
 RA Harris J.L., Sanger F., Naughton M.A.  
 RT "Species differences in insulin."  
 RL Arch. Biochem. Biophys. 65:427-438(1966)  
 RN [2]  
 RP SEQUENCE OF 38-63.  
 RX MEDLINE 73061498; PubMed=4640931;  
 RA Tager H.S., Steiner D.F.  
 RT "Primary structures of the proinsulin connecting peptides of the rat and the horse."  
 RL J. Biol. Chem. 247:7936-7940(1972).  
 CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC -1- CAUTION: X'S AT POSITIONS 31-32 AND 64-65 REPRESENT PAIRED BASIC RESIDUES ASSUMED BY HOMOLOGY TO BE PRESENT IN THE PRECURSOR MOLECULE.

PIR: A01580; IPHQ.  
 HSSP: P01317; IAPH.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism.  
 FT CHAIN 1 30 INSULIN B CHAIN.  
 FT PROPEP 33 63 C PEPTIDE.  
 FT CHAIN 66 86 INSULIN A CHAIN.  
 FT DISULFID 7 72 INTERCHAIN.  
 FT DISULFID 19 85 INTERCHAIN.  
 FT DISULFID 71 76 INTERCHAIN.

SO SEQUENCE 86 AA: 9142 MW: 43E1E8227110R45 CRC64;  
 Query Match 85.1% Score 394; DB 1; Length 86;  
 Best Local Similarity 84.2%; Pred. No. 17e-35;  
 Matches 73; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLYEALVLCGERGFTYTPKTRAEADLOVGGVGLGGPGAGSLQPIALEG 60  
 DB 1 FVNHLCGSHLYEALVLCGERGFTYTPKTRAEADLOVGGVGLGGPGAGSLQPIALEM 60

QY 61 SLOKRGIVEOCCTSCISLYOLENYCN 86  
 DB 61 POUXGIVEOCCTSCISLYOLENYCN 86

## RESULT 9

INS2\_MOUSE STANDARD; PKT; 110 AA.  
 ID INS2\_MOUSE  
 AC P01326;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 13-AUG-1987 (Rel. 05, last sequence update)  
 DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Insulin 2 precursor.  
 GN INS2 OR INS-2.  
 OS Mus musculus (mouse)  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Scleroqgnathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kennedy B.M., Schaffer J.M., Villa-Komaroff L., Chirwin J.M.  
 RT "Characterization of the two pancreatic genes encoding mouse proinsulin."  
 RL J. Mol. Evol. 23:405-412(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NON;  
 RX MEDLINE 90372989; PubMed=2397023;  
 RA Sawa T., Ohnaka S., Morioka H., Yano S.  
 RT "Molecular cloning and DNA sequence analysis of preproinsulin genes in the NON mouse, an animal model of human non-obese, non-insulin-dependent diabetes mellitus."  
 RL J. Mol. Endocrinol. 5:61-67(1990).  
 RN [3]  
 RP SEQUENCE OF 25-54 AND 90-110.  
 RX MEDLINE 72189455; PubMed=5063718;  
 RA Buehli H.F., Glatthaar B., Kunz P., Muelhaupt E., Humbel R.E.  
 RT "Amino acid sequence of the two insulins from mouse (Maus musculus)."  
 RL Hoppe-Seyler's Z. Physiol. Chem. 353:451-458(1972).  
 CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL: X04724; CAA28433.1;  
 PIR: A26342; INMS2  
 HSSP: P01317; IAPH.  
 DR GO: GO:0000187; P:activation of MAPK; IDA.  
 DR GO: GO:0042325; P:regulation of phosphorylation; IDA.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.

INSulin family; Hormone; Glucose metabolism; Signal; Multigene family.  
 SIGNAL 1 24  
 FT CHAIN 25 54 INSULIN 2 B CHAIN.  
 FT PROPEP 57 87 INSULIN 2 C PEPTIDE.  
 FT CHAIN 90 110 INSULIN 2 A CHAIN.  
 FT DISULFID 31 96 INTERCHAIN.

```

FT DISULFID 43 109 INTERCHAIN.
FT DISULFID 95 100
SQ SEQUENCE 110 AA: 12364 MW: 3554C8803D24FDAD CRC64:

Query Match 85.1% Score 394; DB 1; Length 110;
Best Local Similarity 84.9%; Pred. No. 2.2e-35;
Matches 73; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLYEALYLVCGERGFFYTKTREAEDELQVGVQLGGPCAGSLQPIALEG 60
DB 25 FVKQHLCGSHLYEALYLVCGERGFFYTPMSRREVEDPQVAQLGGPCAGSLQPIALEV 84
QY 61 SLOKRGIVQCCTCSLSLYQLENYCN 86
DB 85 AOKRGIVDQCCTCSLSLYQLENYCN 110

RESULT 10
INS2_RAT STANDARD; PRT: 119 AA.
AC P01424
DT 21 JUL 1986 (Rel. 01, Created)
DT 21 JUL 1986 (Rel. 01, Last sequence update)
DE INS2_RAT INS-2
GN INS2_RAT INS-2
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciuror; Muridae; Murinae; Rattus.
OX NCBI_TaxID:10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:Sprague-Dawley; TISSUE:Liver;
RX MEDLINE:80045035; PubMed:498284;
RA Lomedico P., Rosenthal N., Eistratiadis A., Gilbert W., Kolodner R.,
RA Tizard R.;
RT "The structure and evolution of the two nonallelic rat preproinsulin
RT genes.";
RL Cell 18:545-558(1979).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE:86310882; PubMed:2427970;
RA Soares M.B., Schin E., Henderson A., Katathanasis S.K., Cate R.,
RA Zeitlin S., Chirwin J., Eistratiadis A.;
RT "RNA-mediated gene duplication: the rat preproinsulin I gene is a
RT functional retroposon.";
RL Mol. Cell. Biol. 5:2090-2103(1985).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE:80240379; PubMed:6249167;
RA Lomedico P.T., Rosenthal N., Kolodner R., Eistratiadis A.,
RA Gilbert W.;
RT "The structure of rat preproinsulin genes.";
RL Ann. N.Y. Acad. Sci. 343:425-432(1980).
RN [4]
RP SEQUENCE OF 25-54 AND 90-110.
RX MEDLINE:70067613; PubMed:4311918;
RA Steiner D.F., Clark J.L., Nolan C., Rubenstein A.H., Margoliash E.,
RA Aten B., Oyer P.E.;
RT "Proinsulin and the biosynthesis of insulin.";
RL Recent Prog. Horm. Res. 25:207-282(1969).
RN [5]
RP SEQUENCE OF 57-87.
RX MEDLINE:73061498; PubMed:4640931;
RA Tager H.S., Steiner D.F.;
RT "Primary structures of the proinsulin connecting peptides of the rat
RT and the horse.";
RL J. Biol. Chem. 247:7936-7940(1972).
RN [6]
RP SEQUENCE OF 57-87, AND REVISIONS.
RX MEDLINE:72177385; PubMed:4554104;
RA Markussen J., Sundby F.;
RT "Rat-proinsulin C-peptides. Amino-acid sequences.";
```

```

RL Bur. J. Biochem. 25:153-162(1972).
CC FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DB EMBL: V01243; CAA24560.1;
DB EMBL: J00748; AAA41443.1;
DB EMBL: M25585; AAA41442.1;
DB EMBL: M25583; AAA41441.1; JOINED
DB F18; H95789; 18572
DB HISS; P01317; IAPH
DB InterPro: IPR014825; Ins/IGF/relax.
DB Pfam: PF00049; Insulin; 1.
DB SMART: SM00078; IIGF; 1.
DB PROSITE: PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal; Multiqcene family.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN 2 B CHAIN
FT PROPEP 57 87 INSULIN 2 C PEPTIDE.
FT CHAIN 96 110 INSULIN 2 A CHAIN.
FT DISULFID 31 96 INTERCHAIN.
FT DISULFID 43 109 INTERCHAIN.
FT DISULFID 95 100
SQ SEQUENCE 110 AA: 12339 MW: 3A626DA98C86F3CA CRC64:

Query Match 85.1% Score 394; DB 1; Length 110;
Best Local Similarity 84.9%; Pred. No. 2.2e-35;
Matches 73; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLYEALYLVCGERGFFYTKTREAEDELQVGVQLGGPCAGSLQPIALEG 60
DB 25 FVKQHLCGSHLYEALYLVCGERGFFYTPMSRREVEDPQVAQLGGPCAGSLQPIALEV 84
QY 61 SLOKRGIVQCCTCSLSLYQLENYCN 86
DB 85 AOKRGIVDQCCTCSLSLYQLENYCN 110

RESULT 11
INS2_HOIR STANDARD; PRT: 108 AA.
AC P10604
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Insulin precursor.
GN INS.
OS Aotus trivirgatus (Night monkey) (Douroucoulis), and
OS Saimiri sciureus (Common squirrel monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Ceblidae; Aotinae; Aotus.
OX NCBI_TaxID:9505; 9521;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-A. trivirgatus;
RX MEDLINE:88041119; PubMed:3118367;
RA Seino S., Steiner D.F., Bell G.I.;
RT "Sequence of a New World primate insulin having low biological
RT potency and immunoreactivity.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:7423-7427(1987).
```



Cell 18:533-543(1979).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-Sprague-Dawley; TISSUE=Liver;  
 RX MEDLINE=80045035; PubMed=498284;  
 RA Lomedico P., Rosenthal N., Efratiadi A., Gilbert W., Kolodner R.,  
 RT Tizard R.;  
 RA "The structure and evolution of the two nonallelic rat preproinsulin  
 genes";  
 RT Cell 18:545-558(1979).  
 RL  
 RN  
 RP  
 RX  
 RA Lomedico P.T., Rosenthal N., Kolodner R., Efratiadi A.,  
 RA Gilbert W.;  
 RT "The structure of rat preproinsulin genes";  
 RL Ann. N.Y. Acad. Sci. 343:425-432(1980).  
 [4]  
 RN  
 RP  
 RX  
 RA MEDLINE=70067613; PubMed=411948;  
 RA Steiner D.F., Clark J.L., Nollan G., Edelman A.H., Miranlath E.,  
 RA Allen B., Myers P.B.;  
 RT "Proinsulin and the biosynthesis of insulin";  
 RL Revet. Crit. Rev. Biochem. 25:203-282(1989).  
 [5]  
 RN  
 RP  
 RX MEDLINE=7061498; PubMed=4040931;  
 RA Tager H.S., Steiner D.F.;  
 RT "Primary structures of the proinsulin connecting peptides of the rat  
 and the horse";  
 RL J. Biol. Chem. 247:7936-7940(1972).  
 [6]  
 RN  
 RP  
 RX  
 RA MEDLINE=72177385; PubMed=4554104;  
 RA Markussen J., Sundby F.;  
 RT "Rat proinsulin C-peptides. Amino-acid sequences";  
 RL Eur. J. Biochem. 25:153-162(1972).  
 CC  
 CC ! FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC ! SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC ! SURCELLULAR LOCATION: Secreted.  
 CC ! SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC The European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: V01242; CAA24559.1;  
 DR EMBL: J00747; AAA41442.1;  
 DR EMBL: M25584; AAA41439.1;  
 DR PIR: A90788; IPRT1;  
 DR HSP: P01308; IATF;  
 DR InterPro: IPR004825; Ins/IGF/relax  
 DR Pfam: PF00349; Insulin; 1.  
 DR SMART: SM00078; IIGF; 1  
 DR PROSITE: PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabo ism; Signal: Multigene family.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54 INSULIN 1 B CHAIN.  
 FT PROPEP 57 87 INSULIN 1 C PEPTIDE.  
 FT CHAIN 90 110 INSULIN 1 A CHAIN.  
 FT DISULFID 31 96 INTERCHAIN.  
 FT DISULFID 43 109 INTERCHAIN.  
 FT DISULFID 95 100  
 SQ SEQUENCE 110 AA: 12420 MW: 51D606DA54AE3533 CRC64;

Query Match 83.2%; Score 185; DB 1; Length 110;  
 Best Local Similarity 83.7%; Pred. No. 2e-34;  
 Matches 72; Conservative 4; Mismatches 10; Indels 0; Gaps 0;  
 OY 1 FVNQHIGZGSHLEALYLVCGERGFFYTPKTRKAEADLVQGVGGPGAGSLQPLALEG 60  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 25 FVKQHIGZGSHLEALYLVCGERGFFYTPKTRKAEADLVQGVGGPGAGSLQPLALEV 84  
 OY 61 SLOKRGIVEQCCTICSLSLYOLENYCN 86  
 :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 Db 85 ARQKRGIVDCCCTICSLSLYOLENYCN 110  
 RESULT 14  
 INS\_PIG  
 ID INS\_PIG STANDARD; PRT: 138 AA.  
 AC P01315; Q9TSJ5;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DI 16-OCT-2001 (Rel. 40, Last sequence update)  
 DI 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Insulin precursor.  
 GN INS  
 OS Sus scrofa (Pig)  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eularchia; Eularchiiformes; Suidae; Suidae; Sus  
 OX NCBI\_TaxID:9623;  
 RN 1;  
 RP SEQUENCE FROM N.A.  
 RA Han X.G., Tsch B.E.;  
 RT "Complete porcine preproinsulin cDNA sequence";  
 RL Submitted (MAY 1998) to the EMBL/Genbank/DBJ databases.  
 [2]  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN:Large white;  
 RX MEDLINE=22135958; PubMed=12140686;  
 RA Amarger V., Nguyen M., Lueje A.S., Braunschweig M., Nezer C.,  
 RA Georges M., Andersson L.;  
 RT "Comparative sequence analysis of the INS IGF2-H19 gene cluster in  
 RT pigs";  
 RL Mamm. Genome 13:388-398(2002).  
 [3]  
 RN  
 RP SEQUENCE OF 25-108  
 RX MEDLINE=6826486; PubMed=5657063;  
 RA Chance R.E., Ellis R.M., Bromer W.W.;  
 RT "Porcine preinsulin: characterization and amino acid sequence";  
 RL Science 161:165-167(1968).  
 [4]  
 RN  
 RP REVISION TO 59.  
 RA Chance R.E.;  
 RL Submitted (JUL-1970) to the PIR data bank.  
 [5]  
 RN  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RA Blundell T.L., Dodson G.G., Hodgkin D., Mercola D.;  
 RT "Insulin. The structure in the crystal and its reflection in  
 RT chemistry and biology";  
 RL Adv. Protein Chem. 26:279-402(1972).  
 [6]  
 RN  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).  
 RA Isaacs N.W., Agarwal R.C.;  
 RT "Experience with fast Fourier least squares in the refinement of the  
 RT crystal structure of rhombohedral 2-zinc insulin at 1.5-A  
 RT resolution";  
 RL Acta Crystallogr. A 34:782-791(1978).  
 [7]  
 RN  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).  
 RX MEDLINE=89099318; PubMed=2905485;  
 RA Baker E.N., Blundell T.L., Cutfield J.F., Cutfield S.M., Dodson E.J.,  
 RA Dodson G.G., Crowfoot Hodgkin D.M., Hubbard R.E., Isaacs N.W.,  
 RA Reynolds C.D., Sakabe K., Sakabe N., Vijayan N.M.;  
 RT "The structure of 2Zn pig insulin crystals at 1.5-A resolution";  
 RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 319:369-456(1988).  
 [8]  
 RN  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS)

RX MEDLINE-92126280; PubMed-1772633;  
 RA Balschmidt P., Hansen F.B., Dodson E., Dodson G., Korber F.;  
 RT "Structure of porcine insulin cocrystallized with clupeine 2.";  
 RL Acta Crystallogr. B 47:975-986(1991).  
 RN [9]  
 RP X-RAY CRYSTALLOGRAPHY.  
 RX MEDLINE-91222450; PubMed-2025410;  
 RA Badger J., Harris M.R., Reynolds C.D., Evans A.C., Dodson E.J.,  
 RA Dodson G.G., North A.C.T.;  
 RT "Structure of the pig insulin dimer in the cubic crystal";  
 RL Acta Crystallogr. B 47:127-136(1991).  
 RN [10]  
 RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).  
 RA Diao J.-S., Wan Z.-L., Chang W.-R., Liang D. C.;  
 RT "Structure of monomeric porcine DesB1-B2 despentapeptide" (B26-B40)  
 RL Acta Crystallogr. D 53:507-512(1997).  
 CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC -!- DATABASE: NAME-Protein Spotlight.  
 CC NOTE Issue 9 of April 2001:  
 CC WWW="http://www.expasy.org/spotlight/articles/spt1009.html".  
 CC  
 CC This SWISS PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMRL: AF064555; AAC77920.1; ALT\_INIT.  
 DR EMRL: AY044828; AAL69550.1;  
 DR PDB: 4INS; 09-JAN-89.  
 DR PDB: 4INS; 31-JUL-94.  
 DR PDB: 6INS; 31-JAN-94.  
 DR PDB: 7INS; 31-JAN-94.  
 DR PDB: 9INS; 15-OCT-91.  
 DR PDB: 11ZA; 15-OCT-91.  
 DR PDB: 11ZB; 15-OCT-91.  
 DR PDB: 2TCL; 23-JAN-96.  
 DR PDB: 1MPJ; 24-JAN-96.  
 DR PDB: 3MTH; 24-JAN-96.  
 DR PDB: 1DEL; 16-JUN-97.  
 DR PDB: 1SD6; 01-APR-98.  
 DR PDB: 1NAV; 28-FEB-97.  
 DR PDB: 1ZEL; 16-FEB-99.  
 DR PDB: 1ZNI; 28-JAN-98.  
 DR PDB: 1ZNI; 28-JAN-98.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PFC0049; Insulin; 1.  
 DR SMART: SM00078; IGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism; Signal; 3D-structure.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54 INSULIN A CHAIN.  
 FT PROPEP 57 85 C PEPTIDE.  
 FT CHAIN 88 108 INSULIN A CHAIN.  
 FT CHAIN 31 94 INTERCHAIN.  
 FT DISULFID 43 107 INTERCHAIN.  
 FT DISULFID 93 98 INTERCHAIN.  
 FT DISULFID 26 46  
 FT HELIX 48 48  
 FT STRAND 48 48  
 FT HELIX 89 94  
 FT HELIX 100 106  
 FT STRAND 107 107  
 SQ SEQUENCE 108 AA: 11671 MW: CB4491B429858EBE CRC64:

Query Match 82.7%; Score 383; DB 1; Length 108;  
 Best Local Similarity 86.0%; Pred. No. 33e-34;  
 Matches 74; Conservative 1; Mismatches 9; Indels 2; Gaps 1;  
  
 QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRKAEADLVGVGVGQPGGAGLAPLALEG 60  
 DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKTRKAEADLVGVGVGQPGGAGLAPLALEG 82  
 QY 61 SLQKRGIVGQCCISCSLYOLENYCN 86  
 DB 83 PPKRKGIVGQCCISCSLYOLENYCN 108  
  
 RESULT 15  
 ID INS\_PSAOB STANDARD: PRT: 11 AA.  
 AC Q62587;  
 DT 51-NOV-1997 (Ref. 35, 2reated)  
 DT 61-NOV-1997 (Ref. 35, last sequence update)  
 DT 15-JUL-1998 (Ref. 40, last annotation update)  
 DE Insulin, precursor  
 GN INS  
 OS Psammomys obesus.  
 CC Fungiata; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Gerbillinae;  
 CC Psammomys.  
 CC NCBI\_TaxID=48139;  
 RN [1]  
 RP SEQUENCE FROM N A  
 RC TISSUE-Pancreas;  
 RX MEDLINE-97104250; PubMed-9166665;  
 RA Kaiser N., Baillyes E.M., Schneider B.S., Cerasi E., Steiner D.F.,  
 RA Hutton J.C., Gross D.J.;  
 RT "Characterization of the unusual insulin of Psammomys obesus, a  
 RT rodent with nutrition-induced NIDDM-like syndrome";  
 RL Diabetes 46:953-957(1997).  
 CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMRL: X58241; CAA66897.1;  
 DR HSSP: P01308; IAI0  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PFC0049; Insulin; 1.  
 DR PRINTS: PR00277; INSULIN.  
 DR SMART: SM00078; IGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism; Signal.  
 FT SIGNAL 1 24 BY SIMILARITY.  
 FT CHAIN 25 54 INSULIN B CHAIN.  
 FT PROPEP 57 87 C PEPTIDE.  
 FT CHAIN 90 110 INSULIN A CHAIN.  
 FT DISULFID 31 96 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 43 109 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 95 100 BY SIMILARITY.  
 SQ SEQUENCE 110 AA: 12324 MW: A006738E20579CH0 CRC64:

Query Match 81.4%; Score 377; DB 1; Length 110;  
 Best Local Similarity 81.4%; Pred. No. 1.5e-33;

	Matches	70	Conservative	5	Mismatches	11	Indels	0	Gaps	0
Qy	1	FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG	60							
Db	25	FVNOHLCGSHLVEALYLVCGERGFFYTPKFRRGVDDPOMPOLELGGSPGAGDLRALALEV	84							
Qy	61	SIQKRGIVEQCCTGICSLYLENYCN	86							
Db	85	ARQKRGIVEQCCTGICSLYLENYCN	111							

Search completed: September 15, 2003, 12:01:40  
Job time : 9.01434 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:54:30 : Search time 33.5986 Seconds  
(without alignments)  
660,520 Million cell updates/sec

Title: US-09-423-100-4

Perfect score: 463

Sequence: 1 FVNHLCGSHLVKALVLCG.....IVQCCTSLCYLQENYCN 86

Scoring table: BL-SUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 83525

Minimum DB seq length: 0

Maximum DB seq length: 260600000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*

1: sp-archaea:\*

2: sp-bacteria:\*

3: sp-fungi:\*

4: sp-human:\*

5: sp-invertebrate:\*

6: sp-mammal:\*

7: sp-mhc:\*

8: sp-organellae:\*

9: sp-phages:\*

10: sp-plant:\*

11: sp-rodent:\*

12: sp-virus:\*

13: sp-vertebrate:\*

14: sp-unclassified:\*

15: sp-rvivirus:\*

16: sp-bacteriap:\*

17: sp-archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY :

Result No.	Score	Query Match	Length	DB ID	Description
1	463	100.0	110	6	Q8HXV2 ponco pygma
2	413	89.2	110	11	Q91X13 spermophilu
3	408	83.8	110	6	Q8W666 telis silve
4	442	73.9	65	6	Q8Z81 gorilla gor
5	442	73.9	65	6	Q8Z80 ponco pygma
6	246.5	53.2	106	13	Q918Q7 rana pipien
7	235.5	50.9	111	13	Q98TA7 osteoqlssu
8	230.5	49.8	110	13	Q98TA8 pantodon bu
9	222.5	48.1	110	13	Q902Y1 hiodon alos
10	219	47.3	111	13	Q98TB0 chitalla chi
11	214.5	46.3	108	13	Q9DD5 brachydanio
12	212.5	45.9	108	13	Q902N4 catla catla
13	210.5	45.5	87	13	Q98TA9 gnathonemus
14	205.5	44.4	108	13	Q98TB1 catostomus
15	204.5	44.0	91	13	Q98TB2 ambloplites
16	189	40.8	41	11	Q62543 mus spretus

17	162	35.0	39	11	Q62542	mus spretus
18	142.5	30.8	159	13	Q93607	paralichthy
19	142.5	30.8	182	13	Q73720	oreochromis
20	142.5	30.8	182	13	Q42289	oreochromis
21	142.5	30.8	182	13	P79824	oreochromis
22	142.5	30.8	185	13	O57436	oreochromis
23	142.5	30.8	186	13	Q93527	paralichthy
24	141	30.5	207	13	Q90XD0	cyprinus ca
25	140.5	30.3	132	13	Q8AV14	petromyzon
26	138.5	29.9	153	13	Q93380	meleagris g
27	137	29.6	185	13	Q9V157	acanthopag
28	137	29.6	210	13	Q91443	squalus aca
29	136.5	29.5	62	13	Q91AA0	carassius a
30	136.5	29.5	116	13	Q91161	oncorhynch
31	136.5	29.5	117	13	Q91476	salmo salar
32	136.5	29.5	145	13	Q91475	salmo salar
33	136.5	29.5	149	13	Q91231	oncorhynch
34	136.5	29.5	155	13	Q91162	oncorhynch
35	136.5	29.5	161	13	Q91230	oncorhynch
36	136.5	29.5	168	13	P81268	oncorhynch
37	136.5	29.5	168	13	Q91965	oncorhynch
38	135.5	29.3	184	13	Q42336	myoxocephal
39	135.5	29.4	184	13	Q9PSX5	paralichthy
40	135	29.2	212	13	Q8JIE4	brachydanio
41	133.5	28.8	106	6	Q9MY26	trichosurus
42	133.5	28.8	117	13	Q91914	ctenopharyn
43	133.5	28.8	161	13	Q90VV9	brachydanio
44	133.5	28.8	161	13	Q9PKK2	carassius a
45	133.5	28.8	161	13	O98SM6	megalobrama

# ALIGNMENTS

RESULT 1

Q8HXV2 10: Q8HXV2 PRELIMINARY; PRT: 110 AA.  
AC Q8HXV2: 2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)  
UT 01-MAR-2003 (TrEMBLrel. 24, last annotation update)  
DE Insulin precursor.  
GN INS.  
OS Pongo pygmaeus (Orangutan).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
GC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.  
CX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Stead J.D.H., Jeffreys A.J.;  
RT \*Haplotype diversity at the insulin region.\*;  
KL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY37503; AAN06937.1; -;  
SQ SEQUENCE 110 AA; 2202H42H94F520F8 CRC64:

Query Match 100.0%; Score 463; DB 6; Length 110;

Best Local Similarity 100.0%; Pred No 39e-47;

Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNHLCGSHLVKALVLCGGERGFYPTKTRKRAEDLQGVGVLCGSPAGSLQPLALEG 60

.....

Db 25 FVNHLCGSHLVKALVLCGGERGFYPTKTRKRAEDLQGVGVLCGSPAGSLQPLALEG 84

Qy 61 SLOKRGIVFOCTSLCYLQENYCN 86

.....

Db 85 SLOKRGIVFOCTSLCYLQENYCN 110

RESULT 2

Q91X13 10: Q91X13 PRELIMINARY; PRT: 110 AA.

ID Q91X13

AC Q91X13

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

```

DT 01-DHC-2001 (Tremblrel. 19, Last sequence update)
DE Insulin.
OS Sperophilus tridecemlineatus (Thirteen-lined ground squirrel).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
OC Sperophilus.
OX NCBI_TaxID=43179;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Tredrea M.M., Buck M.J., Gubanyiogi J., Squire T.L., Andrews M.T.;
RT "Molecular evolution in a hibernating mammal";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
DR EMBL: AY038604; AAK72558.1;
DR HSSP: P01306; INP.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR0277; INSULIN.
DR SMART: SM0076; IIGF; 1.
DR PROSITE: PS_0262; INSULIN; 1.
SQ SEQUENCE 110 AA: 32004 MW: 45117.686622065 CRC64
Query Match 89.2% Score 414; DB 11; Length 110;
Best Local Similarity 89.5% Pred. No. 3,60-41;
Matches 77; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPAGSLQPLALEG 60
DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKSPREVEDEVGGVGLGGPGAGLPOPLALEM 84

QY 61 SLQKRGIVQCCTSIQSLVQLENYCN 86
DB 85 ALQKRGIVQCCTSIQSLVQLENYCN 110

RESULT 4
Q8H281 PRELIMINARY: PRT: 65 AA.
AC Q8H281;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DI 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DE Insulin (Fragment)
OS Gorilla gorilla (gorilla).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Gorilla.
OX NCBI_TaxID=9593;
RN [1]
RP SEQUENCE FROM N.A.
RA G'dalia G., Tichy H., Klein J.;
RT "Molecular evolution in higher primates: gene specific and organism
specific characteristics";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY092024; AAM76640.1;
DR NCN_TFR 1
DR NON_TFR 65
SQ SEQUENCE 65 AA: 6920 MW: 8772017ED8BCA8EA CRC64;
Query Match 73.9% Score 342; DB 6; Length 65;
Best Local Similarity 100.0% Pred. No. 4,76-83;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPAGSLQPLALEGSLQKRG 66
DB 1 CGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPAGSLQPLALEGSLQKRG 60

QY 67 IVEQC 71
DB 61 IVEQC 65

RESULT 5
Q8H280 PRELIMINARY: PRT: 65 AA.
AC Q8H280;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DI 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DE Insulin (Fragment)
OS Pongo pygmaeus (orangutan).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP SEQUENCE FROM N.A.
RA G'dalia G., Tichy H., Klein J.;
RT "Molecular evolution in higher primates: gene specific and organism
specific characteristics";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY092024; AAM76641.1;
DR NCN_TFR 1
DR NON_TFR 65
SQ SEQUENCE 65 AA: 6920 MW: 8772017ED8BCA8EA CRC64;
Query Match 73.9% Score 342; DB 6; Length 65;
Best Local Similarity 100.0% Pred. No. 4,76-83;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPAGSLQPLALEGSLQKRG 66
DB 1 CGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPAGSLQPLALEGSLQKRG 60

```

```

DT 01-DHC-2001 (Tremblrel. 19, Last sequence update)
DE Insulin.
OS Sperophilus tridecemlineatus (Thirteen-lined ground squirrel).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
OC Sperophilus.
OX NCBI_TaxID=43179;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Tredrea M.M., Buck M.J., Gubanyiogi J., Squire T.L., Andrews M.T.;
RT "Molecular evolution in a hibernating mammal";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
DR EMBL: AY038604; AAK72558.1;
DR HSSP: P01306; INP.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR0277; INSULIN.
DR SMART: SM0076; IIGF; 1.
DR PROSITE: PS_0262; INSULIN; 1.
SQ SEQUENCE 110 AA: 32004 MW: 45117.686622065 CRC64
Query Match 89.2% Score 414; DB 11; Length 110;
Best Local Similarity 89.5% Pred. No. 3,60-41;
Matches 77; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPAGSLQPLALEG 60
DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKSPREVEDEVGGVGLGGPGAGLPOPLALEM 84

QY 61 SLQKRGIVQCCTSIQSLVQLENYCN 86
DB 85 ALQKRGIVQCCTSIQSLVQLENYCN 110

RESULT 4
Q8H281 PRELIMINARY: PRT: 65 AA.
AC Q8H281;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DI 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DE Insulin (Fragment)
OS Gorilla gorilla (gorilla).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Gorilla.
OX NCBI_TaxID=9593;
RN [1]
RP SEQUENCE FROM N.A.
RA G'dalia G., Tichy H., Klein J.;
RT "Molecular evolution in higher primates: gene specific and organism
specific characteristics";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY092024; AAM76640.1;
DR NCN_TFR 1
DR NON_TFR 65
SQ SEQUENCE 65 AA: 6920 MW: 8772017ED8BCA8EA CRC64;
Query Match 73.9% Score 342; DB 6; Length 65;
Best Local Similarity 100.0% Pred. No. 4,76-83;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPAGSLQPLALEGSLQKRG 66
DB 1 CGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPAGSLQPLALEGSLQKRG 60

```



QY 67 IVEQC 71  
 DB 61 IVEQC 65

## RESULT 6

Q91807 PRELIMINARY: PRT: 10 AA.  
 AC Q91807  
 DT 01-JUN-2001 (TREMBLrel. 15, Created)  
 DT 01-JUN-2001 (TREMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Preproinsulin.  
 OS Rana pipiens (Northern leopard frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Ranidae; Rana.  
 OX NCBI\_TaxID=8404;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE 20162607; PubMed=10818274;  
 RA Irwin D.M., Sivarajah P.;  
 RT "Proinsulin cDNAs from the leopard frog, *Rana pipiens*, overlapped  
 RT pretranscription processing."  
 RL Comp. Biochem. Physiol. 125B:405-410(2000).  
 CC -1- SURCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL: AF227187; AAF87285.1;  
 DR HSP: P01315; I5D8.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR PRINTS: PR00277; INSULINB.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 DR NCBI\_TaxID=106 AA; 12183 MW; 3A8709PE/0217F92 CRC64;  
 SQ SEQUENCE 106 AA: 53.2% Score 246.5; DB 13; Length 106;  
 Query Match 53.2% Score 246.5; DB 13; Length 106;  
 Best Local Similarity 51.5% Pred. No. 1.7e-21;  
 Matches 52; Conservative 9; Mismatches 7; Indels 33; Gaps 4;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYPTKTRREAEDLQVQLGGPGAGSLQPL--AL 58  
 DB 24 FDNVLCGSHLVEALYLVCGDGRGFFYPSRRRLD----- --OPLVNGL 65  
 QY 59 FGS-----LQRR--GIVEQCCTSLGSLYLENYCN 96  
 DB 66 QGSELDENVQSQAFQKKKPGIVEQCCHTCTSLYLENYCN 106

## RESULT 7

Q98TA7 PRELIMINARY: PRT: 11; AA.  
 AC Q98TA7  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Preproinsulin (Fragment).  
 OS Osteoglossum bicirrhosum (silver arowana).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
 OC Osteoglossiformes; Osteoglossidae; Osteoglossum.  
 OX NCBI\_TaxID=109271;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-21203577; PubMed=11306171;  
 RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;  
 RT "Molecular cloning of preproinsulin cDNAs from several  
 RT osteoglossomorphs and a cyprinid."  
 RL Mol. Cell. Endocrinol. 174:51-58(2001).  
 CC -1- SURCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL: AF199589; AAK28713.1;  
 DR HSP: P01315; IMPJ.  
 DR InterPro: IPR004825; Ins/IGF/relax.

DR Pfam: PF00049; Insulin; 1.  
 DR PRINTS: PR00277; INSULINB.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 FT NON\_TER 111 111  
 SQ SEQUENCE 111 AA; 12491 MW; AC9F1952D4866D20 CRC64;

Query Match 50.9% Score 235.5; DB 13; Length 111;  
 Best Local Similarity 54.1% Pred. No. 3.7e-20;  
 Matches 46; Conservative 12; Mismatches 26; Indels 1; Gaps 1;  
 QY 3 NOHLCGSHLVEALYLVCGERGFFYPTKTRREAEDLQVQLGGPGAGSLQPLALEGSL 62  
 DB 27 SQHLCGSHLVDALYVMVCGDGRGFFYSPKSRREAEPILGFLSPKSGOENVEVYPKQEGL 86  
 QY 63 Q-KRGIVEQCCTSLGSLYLENYCN 86  
 DB 87 KYKRGIVEQCCHRPQNFIDQNYCN 111

## RESULT 8

Q98TA8 PRELIMINARY: PRT: 110 AA.  
 AC Q98TA8  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Preproinsulin.  
 OS Pantodon buchhoitzi (Butterflyfish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
 OC Osteoglossiformes; Pantodontidae; Pantodon.  
 OX NCBI\_TaxID=8276;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-21203577; PubMed=11306171;  
 RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;  
 RT "Molecular cloning of preproinsulin cDNAs from several  
 RT osteoglossomorphs and a cyprinid."  
 RL Mol. Cell. Endocrinol. 174:51-58(2001).  
 CC -1- SURCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL: AF199588; AAK28712.1;  
 DR HSP: P01308; IHI5.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 DR NCBI\_TaxID=110 AA; 12324 MW; BDECCD659D872E06 CRC64;  
 SQ SEQUENCE 110 AA: 49.8% Score 230.5; DB 13; Length 110;  
 Query Match 49.8% Score 230.5; DB 13; Length 110;  
 Best Local Similarity 46.4% Pred. No. 1.4e-19;  
 Matches 45; Conservative 14; Mismatches 13; Indels 25; Gaps 3;

QY 3 NOHLCGSHLVEALYLVCGERGFFYPTKTRREAEDLQVQLGGPGAGSLQPL----- 58  
 DB 26 SQHLCGSHLVDALYVMVCGDGRGFFYQPKTRVD-----PILGLSPKSAVENE 73  
 QY 59 -----FGSLQ KRGIVEQCCTSLGSLYLENYCN 86  
 DB 74 ADEYYPKMGDLKVRKRGIVEQCCHHPQNFIDQNYCN 110

## RESULT 9

Q90ZY1 PRELIMINARY: PRT: 110 AA.  
 AC Q90ZY1  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Preproinsulin (Fragment).  
 OS Hiodon alosoides (goldeye).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
 OC Osteoglossiformes; Hiodontidae; Hiodon.  
 OX NCBI\_TaxID=54904;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=21203577; PubMed=11306171;  
 RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;  
 RT \*Molecular cloning of preproinsulin cDNAs from several  
 RT osteoglossomorphs and a cyprinid.\*  
 RL Mol. Cell. Endocrinol. 174:51-58(2001).  
 CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL: AF282408; AAK34684.1; -;  
 DR HSSP: P01308; ILPH.

DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR PRINTS: PR00277; INSULINB.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 FT NON\_TER 119 110  
 FT SEQUENCE 119 AA: 1244 MW: 106525.94 GC: 50% CR64;

Query Match 49.1%; Score 214.5; DB 13; Length 110;  
 Best Local Similarity 45.8%; Pred. No. 3.1e-17;  
 Matches 44; Conservative 13; Mismatches 15; Indels 25; Gaps 3;

QY 3 NOHLCGSHLVEALYLVCGERGFFYTPKTRFAEDLVQVQLGQPGAGSLQPLAL;---- 58  
 DB 26 SOHLGSHLVADLYALVCGGKGFYTPKTRFD;-----PLLGFLSPKSAUENE 73

QY 59 -----FCSLQ-KRGIVQVQCTSIQSLYQLENYCN 86  
 DB 74 ADPYPKYQGLKVRGIVEQCCHPCNIFDLNVCN 110

RESULT 10  
 Q98TB0

ID Q98TB0 PRELIMINARY; PRT: 111 AA.  
 AC Q98TH0;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Preproinsulin (Fragment).  
 OS Chitalla chitula (clown knifefish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
 OX NCBI\_TaxID=12163;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21203577; PubMed=11306171;  
 RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;  
 RT \*Molecular cloning of preproinsulin cDNAs from several  
 RT osteoglossomorphs and a cyprinid.\*  
 RL Mol. Cell. Endocrinol. 174:51-58(2001).  
 CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL: AF199586; AAK28710.1; -;  
 DR HSSP: P01308; ILPH.

DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR SMART: SM00078; IIGF; 1.  
 FT NON\_TER 111 111  
 FT SEQUENCE 111 AA: 1248 MW: 247044.43 GC: 42% CR64;

Query Match 47.1%; Score 214.5; DB 13; Length 111;  
 Best Local Similarity 49.0%; Pred. No. 4.3e-18;  
 Matches 48; Conservative 7; Mismatches 17; Indels 26; Gaps 4;

QY 3 NOHLCGSHLVEALYLVCGERGFFYTPKTRFAEDLVQVQLGQPGAGSLQPLAL;LEG 60  
 DB 26 NOHLCGSHLVEALYLVCGERGFFYTPKTRFAEDLVQVQLGQPGAGSLQPLAL;LEG 73

QY 61 SL-----OKRGIVEQCTSIQSLYQLENYCN 86  
 DB 74 EVDFYPPKQGDVMMKRGIVEQCCHPCNIFDLNVCN 111

RESULT 11  
 Q9DD05

ID Q9DD05 PRELIMINARY; PRT: 108 AA.  
 AC Q9DD05;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Insulin precursor.  
 GN INS.  
 OS Brachydanio rerio (zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95425199; PubMed=10495243;  
 RA Arntson F., Zechin F., Hortolunski M.;  
 RT \*Early appearance of pancreatic hormone expressing cells in the  
 RT zebrafish embryo.\*  
 RL Mech. Dev. 87:217-223(1999).  
 CC -1- SURCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL: AJ237750; CAC0109.1; -;  
 DR HSSP: P01308; ILPH.

DR ZFIN: ZDB-GENE-980526-110; Ins  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR PRINTS: PR00277; INSULINB.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.

KW SIGNAL 1 23 POTENTIAL  
 FT CHAIN 24 53 INSULIN B CHAIN.  
 FT CHAIN 86 108 INSULIN A CHAIN.  
 SQ SEQUENCE 108 AA: 11904 MW: 319525.95 GC: 52% CR64;

Query Match 46.3%; Score 214.5; DB 13; Length 108;  
 Best Local Similarity 45.8%; Pred. No. 3.1e-17;  
 Matches 44; Conservative 11; Mismatches 14; Indels 27; Gaps 3;

QY 4 OHLCGSHLVEALYLVCGERGFFYTPKTRFAEDLVQVQLGQPGAGSLQPLAL;EGS-- 61  
 DB 27 OHLCGSHLVADLYLVCGTGFYTPKTRFAEDLVQVQLGQPGAGSLQPLAL;EGS-- 72

QY 62 ----LQKRGIVEQCTSIQSLYQLENYCN 86  
 DB 73 ADFAFKDAELIPKRGIVEQCCHPCNIFDLNVCN 108

RESULT 12  
 Q902N4

ID Q902N4 PRELIMINARY; PRT: 108 AA.  
 AC Q902N4;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Preproinsulin.  
 OS Catla catla (catla).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Catla.  
 OX NCBI\_TaxID=72446;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX Bhattacharya S., Roy S.S., Dasgupta S., Ravikumar L., Mukherjee M.,  
 RA Bandyopadhyaya L., Wakabayashi K.;  
 RT \*A new cell secreting insulin.\*

QY 4 OHLCGSHLVEALYLVCGERGFFYTPKTRFAEDLVQVQLGQPGAGSLQPLAL;EGS-- 61  
 DB 27 OHLCGSHLVADLYLVCGTGFYTPKTRFAEDLVQVQLGQPGAGSLQPLAL;EGS-- 72

QY 62 ----LQKRGIVEQCTSIQSLYQLENYCN 86  
 DB 73 ADFAFKDAELIPKRGIVEQCCHPCNIFDLNVCN 108



Best Local Similarity 46.7%, Pred. No. 1.8e-16;  
Matches 42; Conservative 13; Mismatches 26; Indels 9; Gaps 4;  
QY 4 QHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQ---VGQVELGGGPGA-GSLOPLALE 59  
          |||||:|||||:||||| 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42  
Db 4 QHLCGSHLVDALYLVCGRGFFYNPK--RDVDL--GFLPPKAKGAAAPGGGENEVAEFAFK 61  
          |||||:|||||:||||| 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42  
QY 60 GSLO---KRGIVEQCCTSICSLYOLENYCN 86  
          :: ||||| 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42  
Db 62 DQMEMWVKRGIVEQCCHHPNCNIFDLGRYCN 91  
          :: ||||| 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42

Search completed: September 15, 2003, 12:03:29  
Job time : 35.5986 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:44:15 : Search time 36.5305 Seconds  
(without alignments)  
225.942 Million cell updates/sec

Title: US-09-423-100-5

Perfect score: 294

Sequence: 1 FVNHLCGSHLFALYLVCG.....IVEQCCTICSGLYENYCN 52

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107864 seqs, 154746573 residues

Total number of hits satisfying chosen parameters: 117864

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

1: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL/qcadata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL/qcadata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL/qcadata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSL/qcadata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSL/qcadata/geneseq/geneseq-emb1/AA2003.DAT.\*

Printed No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	294	100.0	52	AA1980	Human insulin prec
2	294	100.0	56	AA1981	Human pro-insulin
3	294	100.0	56	AA1982	Proinsulin sequenc
4	294	100.0	63	AA1983	Human pro-insulin
5	294	100.0	96	AA1984	Human pro-insulin
6	294	100.0	96	AA1985	Fusion protein con
7	294	100.0	107	AA1986	hGH-mini-proinsuli
8	294	100.0	116	AA1987	SOD-proinsulin hyb
9	294	100.0	137	AA1988	Mating factor alph

10	294	100.0	145	16	AA1989	Mating factor alph
11	294	100.0	146	16	AA1990	Mating factor alph
12	294	100.0	150	20	AA1991	Chimeric protein,
13	291	99.9	57	11	AA1992	Proinsulin analog
14	287	97.6	52	12	AA1993	Example of human i
15	283.5	96.4	53	15	AA1994	Di-Arg-(R11-32)-Hu
16	283.5	96.4	53	18	AA1995	Ins double-chain
17	283.5	96.4	53	16	AA1996	Glycosylphosphatid
18	282.5	96.1	160	18	AA1997	SOD-proinsulin hyb
19	282.5	95.7	60	3	AA1998	Human proinsulin a
20	278.5	94.7	51	21	AA1999	Human mature insul
21	278.5	94.7	51	21	AA1999	Human mature insul
22	278.5	94.7	51	23	AA1999	Human insulin, HO
23	278.5	94.7	51	17	AA1999	Modified pJB59 mod
24	278.5	94.7	55	17	AA1999	Modified pJB59 mod
25	278.5	94.7	63	17	AA1999	Single chain insul
26	278.5	94.7	66	17	AA1999	Met-Arg-Met-singl
27	278.5	94.7	87	17	AA1999	Single chain insul
28	278.5	94.7	148	17	AA1999	pK142 modified in
29	277.5	94.4	63	7	AA1999	Sequence of novel
30	277.5	94.4	63	35	AA1999	Gly-A21-di-Arg-(B3
31	277	94.4	65	19	AA1999	Proinsulin 1
32	276.5	94.3	67	18	AA1999	Ins double-chain
33	276.5	94.3	67	18	AA1999	Ins2 double-chain
34	275.5	93.7	53	12	AA1999	Example of human i
35	275.5	93.7	53	17	AA1999	PAK679 modified in
36	275.5	93.7	53	20	AA1999	Modified human ins
37	275.5	93.7	53	22	AA1999	Amino acid sequenc
38	275.5	93.7	55	8	AA1999	Sequence encoded b
39	275.5	93.7	58	17	AA1999	Modified pJB59 mod
40	275.5	93.7	59	17	AA1999	Modified pJB59 mod
41	275.5	93.7	61	17	AA1999	Modified pJB59 mod
42	275.5	93.7	65	17	AA1999	N-terminally exten
43	275.5	93.7	89	17	AA1999	Signal peptide/lea
44	275.5	93.7	91	13	AA1999	Insulin precursor
45	275.5	93.7	91	17	AA1999	YAP3 signal peptid

ALIGNMENTS

RESULT 1  
AA19859  
ID AAY42859 standard: protein; 52 AA.  
AC AAY42859;  
DT 19-JAN 2000 (first entry)  
DE Human insulin precursor, SEQ ID 5.  
XX  
XX Insulin; precursor; growth hormone; chaperone; intramolecular;  
XX folding; conformation; chimeric protein; cleavable; recombinant;  
XX production; yield.  
XX  
XX Homo sapiens.  
XX  
XX W09950302-A1.  
XX  
XX 07-OCT-1999.  
XX  
XX 31-MAR-1998; 98WO-CN00052.  
XX  
XX 31-MAR-1998; 98WO-CN00052.  
XX (TONG) TONGHUA GANTECH BIOTECHNOLOGY LTD.  
XX Gan Z;  
XX WPI: 1999-610839/52.  
XX New chimeric proteins containing human growth hormone fragment, used  
XX particularly for the production of human insulin -  
PT





Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRIGIVEQCCTSIICSLYLENYCN 52  
 DB 45 FVNHLCGSHLVEALYLVCGERGFFYTPKTRIGIVEQCCTSIICSLYLENYCN 96

RESULT 6  
 AAK78662  
 ID AAK78662 standard; protein; 96 AA.  
 AC AAK78662;  
 XX  
 XX  
 XX 03-APR-1996 (first entry)  
 XX  
 XX Fusion protein contg. proinsulin sequence 3.  
 DE  
 XX Proinsulin; post-translational modification; recombinant production;  
 KW protein folding; conformation;  
 KW  
 XX Synthetic;  
 OS  
 XX Key Location/Qualifiers  
 FT Region 41..44  
 FT /label= R2  
 FT /note= "a peptide of 4 amino acids"  
 FT Peptide 45..74  
 FT /label= R1-(R2-R29)-Y  
 FT /note= "human insulin B-chain"  
 FT Region 75  
 FT /label= X  
 FT Peptide 76..96  
 FT /label= Gly (A2-A20)-R1  
 FT /note= "human insulin A-chain"  
 XX  
 PN EPR48292-A2;  
 XX  
 XX 23 AUG-1995;  
 XX  
 XX 09-FEB-1995; 95EP-0101748;  
 XX  
 XX 18 FEB-1994; 94DE-4405179;  
 XX (FARR ) H6ECRST AG;  
 PA  
 PI Gerl M, Ludwig J, Obermeier R, Sabat W;  
 XX WP: 1995-284754/38;  
 DR  
 XX Isolation of insulin that is correctly post translationally  
 PT processed by reacting proinsulin with a mercaptan in the presence  
 PT of a chaotropic agent and purification, after absorption to hydrophobic  
 PT resin  
 XX  
 PS Example 2; Page 8; 16pp; German;  
 CC  
 XX The present sequence is that of a fusion protein, produced in E.coli;  
 CC which contains an example of a proinsulin molecule corresp.  
 CC to the general formula R2-R1-(B2-H29)-Y-X-Gly-(A2-A20)-R3 (17). In  
 CC formula (17), X = Lys, Arg or a peptide of 2-35 amino acids contg.  
 CC Lys or Arg at the N- and C-termini; Y = a natural amino acid; R1 =  
 CC Phe or a bond; R2 = H, Arg, Lys, a peptide of 2-45 amino acids  
 CC contg. Arg or Lys at the N- and C-termini; R3 = a natural amino  
 CC acid; (A2-A20) and (B2-B29) are the insulin A- and B-chain sequences  
 CC from human or other insulin. The proinsulin molecule, released by  
 CC cyanogen bromide, is reacted with mercaptan at a ratio of 2-10 SH  
 CC residues of mercaptan per Cys residue of proinsulin. The reaction  
 CC takes place in the presence of a chaotropic auxiliary agent at  
 CC pH 10-11 and results in proinsulin with correctly linked cysteine  
 CC bridges. Reaction with trypsin and opt. carboxypeptidase B yields  
 CC correctly folded insulin. The insulin is isolated by absorption on  
 CC a hydrophobic resin.  
 XX

SQ Sequence 96 AA;

Query Match 100.0%; Score 294; DB 16; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-26;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRIGIVEQCCTSIICSLYLENYCN 52  
 DB 45 FVNHLCGSHLVEALYLVCGERGFFYTPKTRIGIVEQCCTSIICSLYLENYCN 96

## RESULT 7

AA42860  
 ID AAY42860 standard; protein; 107 AA;  
 XX

AC AAY42860;

XX 19 JAN-2000 (first entry)

XX hGH mini-proinsulin chimeric protein;

KW Insulin precursor; growth hormone; chaperone; intramolecular;  
 KW folding; conformation; chimeric protein; cleaved; recombinant;  
 XX production; yield;  
 XX Synthetic;  
 OS Homo sapiens;  
 OS W09550302-A1;  
 PN 07-01-1995;  
 PD 31-MAR-1998; 98WO-096052;  
 XX 31-MAR-1998; 98WO-096052;  
 XX (TONG-) TONGHUA GENE TECH BIOLOGY LTD.  
 PA Gan Z;

WPI: 1999-610839/52.

XX New chimeric proteins containing human growth hormone fragment, used  
 PT particularly for the production of human insulin;  
 XX

PS Claim 13; Page 30; 45pp; English.

XX This sequence represents a chimeric protein, hGH-mini-proinsulin.  
 CC This chimeric protein contains an N-terminal fragment of human growth  
 CC hormone (hGH) of the sequence given in AAY42855, a cleavable peptide  
 CC linker (AAY42857), and a human insulin precursor comprising insulin  
 CC A and B chains (AAY42859). The hGH portion of the chimeric protein acts  
 CC as an intramolecular chaperone (IMC) for the insulin precursor,  
 CC enabling it to fold correctly. The cleavable peptide linker has a  
 CC C-terminal Arg residue which enables the hGH portion of the  
 CC chimeric protein to be removed after folding has taken place. Production  
 CC of recombinant human insulin via an hGH proinsulin chimeric protein can  
 CC provide human insulin with correctly linked cysteine bridges with  
 CC fewer necessary procedural steps, and hence resulting in a higher yield  
 CC of human insulin. The IMC sequences not only protect insulin sequences  
 CC from intracellular degradation by a microorganism host, but also promote  
 CC the folding of the fused insulin precursor, facilitate the solubility of  
 CC the fusion protein and decrease the intermolecular interactions among  
 CC the fusion proteins, thus allowing folding of the fused insulin precursor  
 CC at commercially useful high concentrations. The procedural steps of  
 CC cyanogen bromide cleavage, oxidative sulphydrololysis and related  
 CC purification steps can thus be eliminated, along with the use of high  
 CC concentrations of mercaptan or the use of hydrophobic absorbent resins.  
 XX

SQ Sequence 107 AA;

Query Match 100.0%; Score 294; DB 20; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-26;



Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRIGIVEQCCTSIQSLYLENYCN 52  
 DB 56 FVNHLCGSHLVEALYLVCGERGFFYTPKTRIGIVEQCCTSIQSLYLENYCN 107

RESULT 8  
 AAR98897  
 ID AAR98897 standard: Protein; 116 AA.  
 XX  
 AC AAR98897;  
 XX  
 DT 01 FEB-1997 (first entry)  
 XX  
 DE SOD proinsulin hybrid polypeptide.  
 XX  
 KW Insulin; proinsulin; hybrid polypeptide; protein folding;  
 KW enzymatic cleavage; cyanogen bromide; saltolysis;  
 XX  
 OS Homo sapiens.  
 XX  
 PN W04620724-A1.  
 XX  
 PD 11 JUL-1996.  
 XX  
 PF 29-DEC-1994; 94WO-US13268.  
 XX  
 PR 29-DEC-1994; 94WO-US13268.  
 PR 10-JAN-1995; 95ZA-0000142.  
 XX  
 PA (BIOT) BIO-TECHNOLOGY GENERAL CORP.  
 XX  
 PI Gorecki M, Hartman JR, Mendelovitz S.  
 XX  
 DR WPI; 1996-331766/33.  
 DR N PSDB; AAR34670.  
 XX  
 PT Recombinant insulin prodn. by correctly folding pro-insulin hybrid  
 PT polypeptide - then enzymatic cleavage of folded product, does not  
 PT require sulphate protection of SH nor use of cyanogen bromide  
 XX  
 PS Example 1H; Figure 7; 69pp; English.  
 XX  
 CC A new method for the production of recombinant human insulin  
 CC comprises folding a hybrid polypeptide comprising proinsulin under  
 CC conditions that permit correct disulphide bond formation and  
 CC subjecting that folded protein to enzymatic cleavage. The insulin  
 CC produced can then be purified. This sequence is a SOD-insulin B  
 CC chain-ArG-insulin A chain hybrid polypeptide and is encoded by  
 CC the plasmid construct pDBAST-LAT. Transformation of the proper  
 CC E.coli host cells with pDBAST-LAT results in the efficient  
 CC expression of the proinsulin hybrid polypeptide, useful for human  
 CC insulin production. The method produces recombinant human insulin  
 CC identical to the natural hormone. Hazardous and cumbersome  
 CC procedures involving cyanogen bromide and sulphitolysis to protect  
 CC SH groups are avoided since the entire hybrid polypeptide folds  
 CC efficiently to the native structure even with the leader attached  
 CC and Cys unprotected.  
 XX  
 SQ Sequence 116 AA;  
 Query Match 100.0%; Score 293; DB 17; Length 116;  
 Best Local Similarity 100.0%; Pred. No. 3,5e-26;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRIGIVEQCCTSIQSLYLENYCN 52  
 DB 65 FVNHLCGSHLVEALYLVCGERGFFYTPKTRIGIVEQCCTSIQSLYLENYCN 116

RESULT 9  
 AAR71692  
 ID AAR71692 standard: Protein; 137 AA.  
 XX  
 AC AAR71692;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 20-NOV-1995 (first entry)  
 XX  
 DE Mating factor alpha 1-Insulin precursor ArqB31.  
 XX  
 KW Human insulin precursor ArqB31; diabetes; zinc ion complex;  
 KW mating factor alpha 1.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Protein 1..85  
 FT /label= mating factor alpha 1  
 FT /peptide 96..116  
 FT /label= B chain  
 FT /peptide 117..137  
 FT /label= A chain  
 XX  
 PN W09507431-A1.  
 XX  
 PD 23 MAR-1995.  
 XX  
 PF 16-SEP-1994; 94WO-DK00147.  
 XX  
 PR 17-SEP-1993; 93DK-0001044.  
 PR 02 FEB-1994; 94OS-0190829.  
 XX  
 PA (NOVO) NOVO-NORDISK AS.  
 XX  
 PI Andersen AS, Walstrom JB, Havelund S, Jonassen L;  
 PI Markussen J;  
 XX  
 DR WPI; 1995-131314/17.  
 DR N-PSDB; AA086425.  
 XX  
 PT Acylated insulin deriv. which may be present as a zinc ion  
 PT complex - is used to treat diabetes and is rapid acting.  
 XX  
 PS Example 5; Page 78; 100pp; English.  
 XX  
 CC AAQ86425 encodes AAR71692 mating factor alpha 1-Insulin precursor  
 CC ArqB31. ArqB31 comprises the B and A chains of a claimed human  
 CC insulin derivative. In the final claimed compsn. they are covalently  
 CC connected via disulphide bonds between Cys residues A7/B7 and  
 CC A20/B19. The derivative, which may be present as a zinc ion  
 CC complex, can be used as a fast action treatment for diabetes.  
 CC (Updated on 25-MAR-2003 to correct PN field.)  
 XX  
 SQ Sequence 137 AA;  
 Query Match 100.0%; Score 294; DB 16; Length 137;  
 Best Local Similarity 100.0%; Pred. No. 4.1e-26;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRIGIVEQCCTSIQSLYLENYCN 52  
 DB 96 FVNHLCGSHLVEALYLVCGERGFFYTPKTRIGIVEQCCTSIQSLYLENYCN 137

RESULT 10  
 AAR71694  
 ID AAR71694 standard: Protein; 145 AA.  
 XX  
 AC AAR71694;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 20-NOV-1995 (first entry)  
 XX  
 DE Mating factor alpha 1-Insulin precursor ArqB1, ArqB31 N-terminal.



```

XX PD 07 OCT-1999.
XX PF 31 MAR-1998; 98WO-CN00052.
XX PR 31 MAR-1998; 98WO-CN00052.
XX PA (TUNG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.
XX PI Gan Z;
XX WPI: 1999-610839/52.
XX New chimeric proteins containing human growth hormone fragment, used
XX particularly for the production of human insulin.
XX Claim 14: Page 30-31; 46pp; English.
XX This sequence represents a chimeric protein, which contains an
XX N-terminal fragment of human growth hormone (hGH) of the sequence given
XX in AA042856, a cleavable peptide linker (AA042857), and a human insulin
XX precursor comprising insulin A and B chains (AA042858). The B-chain portion
XX of the chimeric protein acts as an active insulin chain (1991-09) for
XX the insulin precursor, enabling it to fold correctly into the cleavable
XX peptide linker has a C-terminal Ala residue which makes the helix portion
XX of the chimeric protein to be removed after folding has taken place.
XX production of recombinant human insulin via an hGH-proinsulin chimeric
XX protein can provide human insulin with correctly linked cysteine bridges
XX with fewer necessary procedural steps, and hence resulting in a higher
XX yield of human insulin. The hGH sequences not only protect insulin
XX sequences from intracellular degradation by a microorganism host, but
XX also promote the folding of the fused insulin precursor, facilitate the
XX solubility of the fusion protein and decrease the intermolecular
XX interactions among the fusion proteins, thus allowing folding of the
XX fused insulin precursor at commercially useful high concentrations. The
XX procedural steps of cyanogen bromide cleavage, oxidative sulphatolysis
XX and related purification steps can thus be eliminated, along with the use
XX of high concentrations of mercaptan or the use of hydrophobic absorbent
XX resins.
XX Sequence 150 AA:
SQ
Query Match 100.0%; Score 294; DB 20; Length 150;
Best Local Similarity 100.0%; Pred. No. 4.5e-26;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIICSLYLENYCN 52
Db 99 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIICSLYLENYCN 150

RESULT 14
AA04582
ID AA04582 standard: protein; 57 AA.
XX
XX AA04582;
XX
XX 25-MAR-2003 (updated)
XX 14 SEP 1990 (first entry)
XX Proinsulin analogue with a Lys residue linking the A and B chains.
XX
XX Insulin fusion protein; pro-insulin analogue; tendamistate;
XX Lys Lys bridge; ds
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX misc-difference 36
XX /label=lys residue linking insulin B chain to A chain
XX Peptide 1..35
XX /label= Insulin B chain
XX Peptide 37..57

```

```

FT XX /label= Insulin A chain
FN XX EP367163-A.
XX PD 09-MAY-1990.
XX PF 28-OCT-1989; 89EP-0120056.
XX PR 01-NOV-1988; 88DE-4837273.
XX PR 19-AUG-1989; 89DP-4927449.
XX XX (EAPH ) HOFCHST AG.
XX PI Koller KP, Riess CJ, Uhlmann E, Wallmeier H;
XX WPI: 1990-141149/19.
XX DR N-PSUB: AAQ04335.
XX
XX New insulin fusion proteins comprise pro-insulin analogue linked
XX to tendamistate
XX Disclasured: Page 4; 4pp; German.
XX
XX This sequence is related to the C-terminus of an N-terminal fragment
XX comprising opt. modified tendamistate. This fusion protein
XX may be converted into human insulin using known methods. The synthetic
XX gene was prepared by the phosphoramidite method.
XX See also AA004336.
XX (updated on 25-MAR-2003 to correct PR field.)
XX (updated on 25-MAR-2003 to correct PI field.)
XX SQ Sequence 57 AA:
Query Match 99.0%; Score 291; DB 11; Length 57;
Best Local Similarity 98.1%; Pred. No. 3.8e-26;
Matches 51; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIICSLYLENYCN 52
Db 6 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIICSLYLENYCN 57

RESULT 14
AA011899
ID AA011899 standard: Protein; 52 AA.
XX
XX AA011899;
XX AC
XX 25-MAR 2003 (updated)
XX 22-JUL-1991 (first entry)
XX
XX Example of human insulin precursor.
XX
XX Human insulin; diabetes; transpeptidation.
XX
XX Homo sapiens.
XX
XX EP427296-A.
XX
XX 15-MAY-1991.
XX
XX 24-MAY-1985; 90EP-0121887.
XX
XX 15-NOV-1990; 90EP-0121887.
XX 30-MAY-1984; 84DK-0002665.
XX 08-FEB-1985; 85DK-0000582.
XX
XX (NOVO ) NOVO-NORDISK AS.
XX
XX Markussen J, Fiil N, Ammerer G, Hansen MT, Thim L;
XX Norris K, Volqt RO;
XX WPI: 1991-141828/20.

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:57:40 : Search time 11 9283 Seconds  
( without alignments)  
184.449 Million cell updates/sec

Title: US-09-423-100-5

Perfect score: 294

Sequence: 1 FVNHLCGSHLVEALYLVCG.....IVEQCCTISCSLYOLENYCN 52

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 428717 seqs, 4241658 residues

Total number of hits satisfying chosen parameters 12477

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA\*  
1: /cqn2\_6/plodata/1/1aa/5A... .pep:\*  
2: /cqn2\_6/plodata/1/1aa/5B... .pep:\*  
3: /cqn2\_6/plodata/1/1aa/6A... .pep:\*  
4: /cqn2\_6/plodata/1/1aa/6B... .pep:\*  
5: /cqn2\_6/plodata/1/1aa/PC10S... .pep:\*  
6: /cqn2\_6/plodata/1/1aa/backfiles... .pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	294	100.0	56	1	US-08-160-376A-7 Sequence 7, Appl
2	294	100.0	56	1	US-08-389-487-11 Sequence 11, Appl
3	294	100.0	63	1	US-08-160-376A-6 Sequence 6, Appl
4	294	100.0	66	1	US-08-291-060B-5 Sequence 5, Appl
5	294	100.0	96	1	US-08-160-376A-5 Sequence 5, Appl
6	294	100.0	96	1	US-08-389-487-8 Sequence 8, Appl
7	294	100.0	137	1	US-08-400-256-39 Sequence 39, Appl
8	294	100.0	137	3	US-08-975-365-39 Sequence 39, Appl
9	294	100.0	145	1	US-08-400-256-45 Sequence 45, Appl
10	294	100.0	145	3	US-08-975-365-45 Sequence 45, Appl
11	294	100.0	146	1	US-08-400-256-48 Sequence 48, Appl
12	294	100.0	146	3	US-08-975-365-48 Sequence 48, Appl
13	291	99.0	57	1	US-08-030-731A-44 Sequence 44, Appl
14	283.5	96.4	53	1	US-08-233-617-4 Sequence 4, Appl
15	283.5	96.4	53	4	US-08-981-988A-42 Sequence 42, Appl
16	278.5	94.7	51	4	US-09-472-948-3 Sequence 3, Appl
17	278.5	94.7	51	4	US-09-723-981-3 Sequence 3, Appl
18	277.5	94.7	51	4	US-09-723-896-3 Sequence 3, Appl
19	277.5	94.4	53	1	US-08-233-617-3 Sequence 3, Appl
20	277.5	94.2	65	3	US-08-900-574-3 Sequence 3, Appl
21	276.5	94.0	55	3	US-08-900-574-6 Sequence 6, Appl
22	276.5	94.0	66	3	US-08-900-574-5 Sequence 5, Appl
23	276.5	94.0	67	4	US-08-981-988A-1 Sequence 1, Appl
24	276.5	94.0	67	4	US-08-981-988A-5 Sequence 5, Appl
25	276.5	93.9	67	3	US-08-901-853-7 Sequence 7, Appl
26	275.5	93.7	53	3	US-09-261-873-2 Sequence 2, Appl
27	275.5	93.7	65	1	US-08-468-674B-71 Sequence 71, Appl

Sequence 71, Appl  
Sequence 41, Appl  
Sequence 41, Appl  
Sequence 45, Appl  
Sequence 15, Appl  
Sequence 15, Appl  
Sequence 37, Appl  
Sequence 3, Appl  
Sequence 36, Appl  
Sequence 19, Appl  
Sequence 33, Appl  
Sequence 42, Appl  
Sequence 33, Appl  
Sequence 42, Appl  
Sequence 2, Appl  
Sequence 4, Appl

## ALIGNMENTS

RESULT :  
US-08-160-376A-7  
: Sequence 7, Application US/08160376A  
: Patent No. 5473049  
: GENERAL INFORMATION:  
: APPLICANT: Obermeier, Rarier  
: APPLICANT: Gerl, Martin  
: APPLICANT: Ludwig, Jurgen  
: APPLICANT: Sabat, Walter  
: TITLE OF INVENTION: Process For Obtaining Proinsulin  
: TITLE OF INVENTION: Possessing Correctly Linked  
: TITLE OF INVENTION: Cystine Bridges  
: NUMBER OF SEQUENCES: 7  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Kenneth A. Genoni, Esq.  
: STREET: Rt. 202-06 No. 5473049th/P.O. Box 2500  
: CITY: Somerville  
: STATE: New Jersey  
: COUNTRY: U.S.A.  
: ZIP: 08876-1258  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
: COMPUTER: IBM 386  
: OPERATING SYSTEM: WINDOWS 3.1  
: SOFTWARE: WORDPERFECT 5.1  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/160.376A  
: FILING DATE: December 1, 1993  
: CLASSIFICATION: C10  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: GE P 4240420-7  
: FILING DATE: December 2, 1992  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Barbara V. Maurer, Esq.  
: REGISTRATION NUMBER: 31,287  
: REFERENCE/DOCKET NUMBER: HOE 92/F 384  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (908) 231-4079  
: TELEFAX: (908) 231-2255  
: INFORMATION FOR SEQ ID NO: 7:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 56 Amino Acids  
: TYPE: Amino Acid (AA)  
: TOPOLOGY: not relevant  
US-08-160-376A-7

Query Match 100.0% Score 244: DB 1: Length 56:  
Best Local Similarity 100.0% Pred. No. 4.9e-29;  
Matches 52: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTICSILYOLENYCN 52  
DB 5 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTICSILYOLENYCN 56

## RESULT 2

US-08-389-487-11  
Sequence 11, Application US/08389487  
Patent No. 5663291  
GENERAL INFORMATION:  
APPLICANT: Obermeyer, Rainer  
APPLICANT: Gerl, Martin  
APPLICANT: Ludwig, Jurgen  
APPLICANT: Sabell, Walter  
TITLE OF INVENTION: Process for Obtaining Insulin Having  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
STREET: 1400 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/389,487  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Einaudi, Carol P.  
REGISTRATION NUMBER: 32,220  
REFERENCE/POCKET NUMBER: 02481.1424-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 56 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
MOLECULE TYPE: peptide  
US-08-389-487 11

Query Match 100.0% Score 244, DB 1, Length 56:  
Best Local Similarity 100.0% Pred. No. 4, 9e+25;  
Matches 52: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTICSILYOLENYCN 52  
DB 5 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTICSILYOLENYCN 56

## RESULT 3

US-08-160-376A-6  
Sequence 6, Application US/08160376A  
Patent No. 5473049  
GENERAL INFORMATION:  
APPLICANT: Obermeyer, Rainer  
APPLICANT: Gerl, Martin  
APPLICANT: Ludwig, Jurgen  
APPLICANT: Sabell, Walter  
TITLE OF INVENTION: Process For Obtaining Proinsulin  
TITLE OF INVENTION: Possessing Correctly  
TITLE OF INVENTION: Cystine Bridges  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Kenneth A. Genoni, Esq.  
STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500  
CITY: Somerville  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 08876-1268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM 386  
OPERATING SYSTEM: WINDOWS 3.1  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/160,376A  
FILING DATE: December 1, 1993  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GE P 4240420.7  
FILING DATE: December 2, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Barbara V. Mueller, Esq.  
REGISTRATION NUMBER: 31,287  
REFERENCE/POCKET NUMBER: HIR 92/F 444  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 231-4079  
TELEFAX: (908) 231-2255  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 63 Amino Acids  
TYPE: Amino Acid (AA)  
TOPOLGY: not relevant  
US-08-160-376A-6

Query Match 100.0% Score 244, DB 1, Length 63:  
Best Local Similarity 100.0% Pred. No. 5, 6e+25;  
Matches 52: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTICSILYOLENYCN 52  
DB 12 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTICSILYOLENYCN 63

## RESULT 4

US-08-291-060B-5  
Sequence 5, Application US/08291060B  
Patent No. 5728543  
GENERAL INFORMATION:  
APPLICANT: Dorschug, Michael  
APPLICANT: Kolier, Klaus-Peter  
APPLICANT: Marquardt, Rudiger  
APPLICANT: Melwes, Johannes  
TITLE OF INVENTION: An Enzymatic Process for the  
TITLE OF INVENTION: Conversion of Preproinsulins Into Insulins  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/291,060B  
FILING DATE: 08-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Einaudi, Carol P.  
REGISTRATION NUMBER: 32,220

```

? REFERENCE/DOCKET NUMBER: 02481.1105-0-0000
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 408-4366
? TELEFAX: (202) 408-4400
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 66 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: peptide
? US-08-291-0608-5
?
? Query Match 100.0%; Score 294; DB 1; Length 66;
? Best Local Similarity 100.0%; Pred. No. 5.9e-29;
? Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
?
? QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 52
? DB 15 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 66
?
? RESULT 5
? US-08-160-376A-5
? Sequence 5, Application US/08160376A
? Patent No. 5471049
? GENERAL INFORMATION:
? APPLICANT: Obermeier, Rainer
? APPLICANT: Gerl, Martin
? APPLICANT: Ludwig, Jurgen
? APPLICANT: Sabel, Walter
? TITLE OF INVENTION: Process for obtaining proteins
? TITLE OF INVENTION: Possessing correctly linked
? NUMBER OF SEQUENCES: 7
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Kenneth A. Geroni, Esq
? STREET: Rt. 202-206 No. 5473049H/P.O. Box 2500
? CITY: Somerville
? STATE: New Jersey
? COUNTRY: U.S.A.
? ZIP: 08876-1258
? COMPUTER READABLE FORM:
? MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
? COMPUTER: IBM 386
? OPERATING SYSTEM: WINDOWS 3.1
? SOFTWARE: WORDPERFECT 5.1
? CURRENT APPLICATION DATA:
? FILING DATE: December 1, 1993
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: GE P 4240420.7
? FILING DATE: December 2, 1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Barbara V. Maurer, Esq
? REGISTRATION NUMBER: 31,287
? REFERENCE/DOCKET NUMBER: H/E 92/P 364
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (908) 231-4079
? TELEFAX: (908) 231-2255
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 66 Amino Acids
? TYPE: Amino Acid (AA)
? TOPOLOGY: not relevant
? US-08-160-376A-5
?
? Query Match 100.0%; Score 294; DB 1; Length 96;
? Best Local Similarity 100.0%; Pred. No. 4.8e-29;
? Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
?
? QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 52
? DB 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 96
?
? RESULT 6
? US-08-389-487-8
? Sequence 8, Application US/08389487
? Patent No. 5663291
? GENERAL INFORMATION:
? APPLICANT: Obermeier, Rainer
? APPLICANT: Gerl, Martin
? APPLICANT: Ludwig, Jurgen
? APPLICANT: Sabel, Walter
? TITLE OF INVENTION: Process for obtaining Insulin Having
? TITLE OF INVENTION: Correctly linked Cysteine Bridges
? NUMBER OF SEQUENCES: 12
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
? ADDRESSEE: Eganer
? STREET: 1400 I Street, N.W.
? CITY: Washington
? STATE: D.C.
? COUNTRY: United States of America
? ZIP: 20005-3415
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/189,487
? FILING DATE:
? CLASSIFICATION: 530
? ATTORNEY/AGENT INFORMATION:
? NAME: Einaudi, Carol P.
? REGISTRATION NUMBER: 32,119
? REFERENCE/DOCKET NUMBER: 7481.1424-000000
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-408-4000
? TELEFAX: 202-408-4400
? INFORMATION FOR SEQ ID NO: 8:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 96 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: peptide
? US-08-389-487-8
?
? Query Match 100.0%; Score 294; DB 1; Length 96;
? Best Local Similarity 100.0%; Pred. No. 8.8e-29;
? Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
?
? QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 52
? DB 45 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 96
?
? RESULT 7
? US-08-400-256-39
? Sequence 19, Application US/08400256
? Patent No. 5750497
? GENERAL INFORMATION:
? APPLICANT: Havelund, Svend
? APPLICANT: Halstrom, John
? APPLICANT: Jonassen, Ib
? APPLICANT: Andersen, Asger Sloth
? APPLICANT: Markussen, Jan
? TITLE OF INVENTION: ACYLATED INSULIN
? NUMBER OF SEQUENCES: 49
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: No. 5750497a No. 5750497b disk of No. 5750497th America, Inc.
? STREET: 405 Lexington Avenue, 64th Floor

```

CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/400,256  
FILING DATE: 03-MAR-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 3985, 220-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 45  
SEQUENCE CHARACTERISTICS:  
LENGTH: 137 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-400-256-39

Query Match 100.0%; Score 254; DB 1; Length 137;  
Best Local Similarity 100.0%; Pred. No. 1 3e-28;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVGFCCTSIQSLYLENYCN 52  
DB 86 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRG VEQCTSIQSLYLENYCN 137

RESULT 8  
US-08-975-365-39  
Sequence 39, Application US/08975365  
Patent No. 6011007  
GENERAL INFORMATION:  
APPLICANT: Havelund, Svend  
APPLICANT: Halstrom, John  
APPLICANT: Jonassen, Ib  
APPLICANT: Andersen, Asger Sloth  
APPLICANT: Markussen, Jan  
TITLE OF INVENTION: ACYLATED INSULIN  
NUMBER OF SEQUENCES: 49  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6011007a No. 6011007disk of No. 6011007th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/975,365  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/400,256  
FILING DATE: 03-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 3985, 220-US  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 39;  
SEQUENCE CHARACTERISTICS:  
LENGTH: 137 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-975-365-39

Query Match 100.0%; Score 294; DB 4; Length 137;  
Best Local Similarity 100.0%; Pred. No. 1 3e-28;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVGFCCTSIQSLYLENYCN 52  
DB 86 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVGFCCTSIQSLYLENYCN 137

RESULT 9  
US-08-400-256-45  
Sequence 45, Application US/08400256  
Patent No. 5750497  
GENERAL INFORMATION:  
APPLICANT: Havelund, Svend  
APPLICANT: Halstrom, John  
APPLICANT: Jonassen, Ib  
APPLICANT: Andersen, Asger Sloth  
APPLICANT: Markussen, Jan  
TITLE OF INVENTION: ACYLATED INSULIN  
NUMBER OF SEQUENCES: 49  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5750497a No. 5750497disk of No. 5750497th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/400,256  
FILING DATE: 03-MAR-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 3985, 220-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 45;  
SEQUENCE CHARACTERISTICS:  
LENGTH: 145 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-400-256-45

Query Match 100.0%; Score 294; DB 3; Length 145;  
Best Local Similarity 100.0%; Pred. No. 1 4e-28;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVGFCCTSIQSLYLENYCN 52  
DB 94 FVNOHLCGSHLVEALYLVCG :RGFFYTPKTRGIVGFCCTSIQSLYLENYCN 145

RESULT 10  
US-08-975-365-45



```

Sequence 45, Application US/08975365
Patent No. 6011007
GENERAL INFORMATION:
APPLICANT: Havelund, Svend
APPLICANT: Halstrom, John
APPLICANT: Jonassen, Ib
APPLICANT: Andersen, Asger Sloth
APPLICANT: Markussen, Jan
TITLE OF INVENTION: ACYLATED INSULIN
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6011007o No. 6011007disk of No. 6011007th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,365
FILING DATE: 03-MAR-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/400,256
FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3985,220-05
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 145 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-975-365-45

Query Match 100.0% Score 244: DB 3: Length 145:
Best Local Similarity 100.0% Pred. No. 1,4e-28:
Matches 52: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 1 FVNQHCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCVTSICSLYQLENYCN 52
DB 94 FVNQHCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCVTSICSLYQLENYCN 145

RESULT 11
US-08-400-256-48
Sequence 48, Application US/08400256
Patent No. 5750497
GENERAL INFORMATION:
APPLICANT: Havelund, Svend
APPLICANT: Halstrom, John
APPLICANT: Jonassen, Ib
APPLICANT: Andersen, Asger Sloth
APPLICANT: Markussen, Jan
TITLE OF INVENTION: ACYLATED INSULIN
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5750497o No. 5750497disk of No. 5750497th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,365
FILING DATE: 03-MAR-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/400,256
FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3985,220-05
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 145 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-975-365-48

Query Match 100.0% Score 244: DB 1: Length 146:
Best Local Similarity 100.0% Pred. No. 1,4e-28:
Matches 52: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 1 FVNQHCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCVTSICSLYQLENYCN 52
DB 95 FVNQHCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCVTSICSLYQLENYCN 146

RESULT 12
US-08-975-365-48
Sequence 48, Application US/08975365
Patent No. 6011007
GENERAL INFORMATION:
APPLICANT: Havelund, Svend
APPLICANT: Halstrom, John
APPLICANT: Jonassen, Ib
APPLICANT: Andersen, Asger Sloth
APPLICANT: Markussen, Jan
TITLE OF INVENTION: ACYLATED INSULIN
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6011007o No. 6011007disk of No. 6011007th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,365
FILING DATE: 03-MAR-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/400,256
FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3985,220-05
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 146 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-400-256-48

```

? TYPE: amino acid  
? TOPOLOGY: linear  
? MOLECULE TYPE: protein  
US-08-975-365-48

Query Match 100.0%; Score 294; DB 3; Length 146;  
Best Local Similarity 100.0%; Pred. No. 1.4e-28;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNOHLCGSHLVGCGRGGFFTPKTRGIVEOCCTSCISLYOLENYCN 52  
DB 95 FVNOHLCGSHLVGCGRGGFFTPKTRGIVEOCCTSCISLYOLENYCN 145  
|||||

RESULT 13  
US-08-030-731A-44  
? Sequence 44, Application US/08010731A  
? Patent No. 5426046  
? GENERAL INFORMATION:  
? APPLICANT: Kollon, Klaus Peter  
? APPLICANT: Kress, Guenter Hubertus  
? APPLICANT: Lehmann, Eugen  
? APPLICANT: Wallmeier, Wolfram  
? TITLE OF INVENTION: Processes for the Preparation of Foreign  
? TITLE OF INVENTION: Proteins in Streptomyces  
? NUMBER OF SEQUENCES: 48  
? CORRESPONDENCE ADDRESS:  
? ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
? ADDRESSEE: Dunner  
? STREET: 1300 I Street, N.W., Suite 700  
? CITY: Washington  
? STATE: D.C.  
? COUNTRY: USA  
? ZIP: 20005-4315  
? COMPUTER READABLE FORM:  
? MEDIUM TYPE: Floppy disk  
? COMPUTER: IBM PC compatible  
? OPERATING SYSTEM: PC-DOS/MS-DOS  
? SOFTWARE: Patent In Release #1.0, Version #1.25  
? CURRENT APPLICATION DATA:  
? APPLICATION NUMBER: US/08/040,741A  
? FILING DATE: 12-MAR-1993  
? CLASSIFICATION: 435  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: US 07/189,840  
? FILING DATE: 03-MAY-1988  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: US 07/430,622  
? FILING DATE: 01-NOV-1989  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: US 07/687,610  
? FILING DATE: 19-APR-1991  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: US 07/735,757  
? FILING DATE: 29-JUL-1991  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: DE P 37 14 866.4  
? FILING DATE: 05-MAY-1987  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: DE P 48 37 273.4  
? FILING DATE: 03-NOV-1988  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: DE P 19 27 445.7  
? FILING DATE: 19-AUG-1989  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: DE P 40 12 816.0  
? FILING DATE: 21-APR-1990  
? ATTORNEY/AGENT INFORMATION:  
? NAME: Kirschner Michael K.  
? REGISTRATION NUMBER: 34,851  
? REFERENCE/DOCKET NUMBER: 02481-0593-02000  
? TELECOMMUNICATION INFORMATION:  
? TELEPHONE: 202-408-4000

? TELEFAX: 202-408-4400  
? INFORMATION FOR SEQ ID NO: 44:  
? SEQUENCE CHARACTERISTICS:  
? LENGTH: 57 amino acids  
? TYPE: amino acid  
? TOPOLOGY: unknown  
? MOLECULE TYPE: peptide  
US-08-030-731A-44

Query Match 99.0%; Score 291; DB 1; Length 57;  
Best Local Similarity 98.1%; Pred. No. 1.2e-28;  
Matches 51; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FVNOHLCGSHLVGCGRGGFFTPKTRGIVEOCCTSCISLYOLENYCN 52  
DB 6 FVNOHLCGSHLVGCGRGGFFTPKTRGIVEOCCTSCISLYOLENYCN 57  
|||||

RESULT 14  
US-08-233-617-4  
? Sequence 1, Application US/080233617  
? Patent No. 5466066  
? GENERAL INFORMATION:  
? APPLICANT: Obermeyer, Kainer  
? APPLICANT: Sabel, Walter  
? APPLICANT: Deil, Peter  
? APPLICANT: Geisen, Ralf  
? TITLE OF INVENTION: Amphiphous Microspherical Forms of Insulin  
? TITLE OF INVENTION: Derivatives  
? NUMBER OF SEQUENCES: 4  
? CORRESPONDENCE ADDRESS:  
? ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
? ADDRESSEE: Dunner  
? STREET: 1300 I Street, N.W., Suite 700  
? CITY: Washington  
? STATE: D.C.  
? COUNTRY: USA  
? ZIP: 20005-4315  
? COMPUTER READABLE FORM:  
? MEDIUM TYPE: Floppy disk  
? COMPUTER: IBM PC compatible  
? OPERATING SYSTEM: PC-DOS/MS-DOS  
? SOFTWARE: Patent In Release #1.0, Version #1.25  
? CURRENT APPLICATION DATA:  
? APPLICATION NUMBER: US/08/233,617  
? FILING DATE: 25-APR-1994  
? CLASSIFICATION: 514  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: P 43 13 752.4  
? FILING DATE: 27-APR-1993  
? ATTORNEY/AGENT INFORMATION:  
? NAME: Carol P. Einaudi  
? REGISTRATION NUMBER: 32,220  
? REFERENCE/DOCKET NUMBER: 02481.1374 00000  
? TELECOMMUNICATION INFORMATION:  
? TELEPHONE: 202 408-4000  
? TELEFAX: 202-408-4400  
? INFORMATION FOR SEQ ID NO: 4:  
? SEQUENCE CHARACTERISTICS:  
? LENGTH: 53 amino acids  
? TYPE: amino acid  
? TOPOLOGY: linear  
? MOLECULE TYPE: protein  
? ORIGINAL SOURCE:  
? ORGANISM: Escherichia coli  
US-08-233-617-4

Query Match 96.4%; Score 283.5; DB 1; Length 53;  
Best Local Similarity 98.1%; Pred. No. 8.8e-28;  
Matches 52; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 FVNOHLCGSHLVGCGRGGFFTPKTRGIVEOCCTSCISLYOLENYCN 52  
|||||

Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRRGIVEQCCTSI CSLYQLENYCN 53

## RESULT IS

US-08-981-988A-42

; Sequence 42, Application US/08981988A

; Patent No. 6337194

GENERAL INFORMATION:

APPLICANT: Vittal Mallya Scientific Research Foundation

**APPLICANT:** The University of Leicester

TITLE OF INVENTION: Insulin

NUMBER OF SEQUENCES: 43

CORRESPONDENCE ADDRESS:

ADDRESSEE: VITTAL MALLA SCIENTIFIC RESEARCH FOUNDATION

STREET: K. R. ROAD

CITY: BANGALORE

COUNTRY: INDIA

CONFIDENTIAL: INDI  
2000 560 1014

COMPTON RESEARCH FORM

THE UNIVERSITY OF CHICAGO

REDACTION, INC., 100 BY GLS  
S. MOORE, 100 BY GLS

COMPLERE IBM PC COMPALGLO  
CORPATING SYSTEM PC-1000/10000

[illegible]

**SUPPLIERS:** Patent in Re...

; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: 115 000 000 000

APPLICATION I

FILING DATE: 11/11/2011

; CLASSIFICATION: 435

PRIME APPLICATION DATA:

APPLICATION NUMBER: GR 9

FILING DATE: 08-JUL-1995

; INFORMATION FOR SEQ ID NO:

; SEQUENTIAL CHARACTERISTICS:

; LENGTH: 53 amino

TYPE: amino

; STRANDEDNESS:

: ; TOPOLCAS:

### Query Match

Query Match	96.48	Score	283	DB 4	Length	53
-------------	-------	-------	-----	------	--------	----

Query Index	Score	Pred.
Best local similarity	98.18	98.28

1 EVNQHICCSHLYEALVLCGERGEFFYTPKT-RQIVFQCCITSCSYVQI ENYCN 52

QY I FVNQHLGGSALVEXATL EVCGERGFF I PKI -RGIVEGQCISGSGS::IQENICN 32

Search completed: September 15, 2003. 12:05:32

Job Time : 11.9283 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 12:03:35 ; Search time 21.8065 Seconds  
(without alignments)  
347.945 Million cell updates/sec

Title: US-09-423-100-5

Perfect score: 294

Sequence: 1 FVNHGSHLVEALYLVCGL...IVEQCCTICSLYQLNVCN 52

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 543946 seqs, 14591246 residues

Total number of hits satisfying chosen parameters: 541946

Minimum DB seq length: 6

Maximum hit seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCI\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY:

Result No.	Score	Query Match	Length	DB ID	Description
1	294	100.0	52	14	US-10-054-873-5
2	294	100.0	107	14	US-10-054-873-6
3	294	100.0	150	14	US-10-054-873-7
4	278.5	94.7	51	11	US-09-878-978-5
5	278.5	94.7	51	14	US-10-028-410-1
6	275.5	93.7	124	10	US-09-894-711-18
7	275.5	93.7	138	10	US-09-861-687-19
8	273	92.9	50	14	US-10-066-0094-3
9	267	90.8	86	10	US-09-878-380-1
10	267	90.8	86	11	US-09-878-978B-4
11	267	90.8	86	14	US-10-028-410-2
12	267	90.8	86	14	US-10-054-873-4
13	267	90.8	96	10	US-09-947-563-4
14	267	90.8	110	9	US-09-205-658-125
15	267	90.8	110	9	US-09-815-229-3

16	267	90.8	110	10	US-09-804-409A-9	Sequence 9, Appli
17	267	90.8	110	12	US-09-969-748C-6	Sequence 6, Appli
18	267	90.8	110	15	US-10-038-686-1	Sequence 1, Appli
19	267	90.8	110	15	US-10-028-813-2	Sequence 2, Appli
20	267	90.8	117	9	US-09-280-030-63	Sequence 63, Appli
21	267	90.8	130	9	US-09-280-030-62	Sequence 62, Appli
22	266.5	90.6	124	9	US-09-736-611-12	Sequence 12, Appli
23	266.5	90.6	124	9	US-09-740-359-12	Sequence 12, Appli
24	266.5	90.6	124	10	US-09-894-711-12	Sequence 12, Appli
25	266.5	90.6	125	9	US-09-736-611-10	Sequence 10, Appli
26	266.5	90.6	125	9	US-09-740-359-10	Sequence 10, Appli
27	266.5	90.6	125	10	US-09-894-711-10	Sequence 10, Appli
28	266.5	90.6	147	9	US-09-736-611-8	Sequence 8, Appli
29	266.5	90.6	147	9	US-09-740-359-7	Sequence 7, Appli
30	265	90.1	144	9	US-09-736-611-6	Sequence 6, Appli
31	265	90.1	144	9	US-09-740-359-5	Sequence 5, Appli
32	265	90.1	146	10	US-09-894-711-5	Sequence 5, Appli
33	261	88.9	96	10	US-09-947-563-5	Sequence 5, Appli
34	244.5	83.2	144	10	US-09-894-711-7	Sequence 7, Appli
35	219	74.5	46	9	US-09-205-658-132	Sequence 132, App
36	214	72.8	46	9	US-09-205-658-133	Sequence 133, App
37	210	68.0	46	9	US-09-205-658-135	Sequence 135, App
38	195	66.3	46	9	US-09-205-658-134	Sequence 134, App
39	194	66.0	46	9	US-09-205-658-136	Sequence 136, App
40	184	62.6	46	9	US-09-205-658-137	Sequence 137, App
41	183.5	62.4	50	9	US-09-205-658-115	Sequence 115, App
42	173	58.8	32	9	US-09-815-229-14	Sequence 14, Appli
43	173	58.8	32	10	US-09-947-563-6	Sequence 6, Appli
44	173	58.8	56	9	US-09-815-229-10	Sequence 10, Appli
45	173	58.8	166	9	US-09-925-297-805	Sequence 805, App

# ALIGNMENTS

RESULT 1  
US-10-054-873-5  
; Sequence 5, Application US/10054873  
; Publication No. US20020164712A1  
GENERAL INFORMATION:  
; APPLICANT: Gan, Zhong Ru  
; TITLE OF INVENTION: Chimeric Protein Containing an Intramolecular Chaperone Like Sequence  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3844  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/054-873  
; FILING DATE: 22-Jan-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/CN98/00052  
; FILING DATE: 31-MAR-1998  
; APPLICATION NUMBER: US 09/423,100  
; FILING DATE: 11-DEC-2000  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mycroft, Frank J  
; REGISTRATION NUMBER: 46,946  
; REFERENCE/DOCKET NUMBER: 020167-00613005  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 52 amino acids  
; TYPE: amino acid



```
Matches 51: Conservative 0: Mismatches 0: Indels 1: Gaps 1:
Oy 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 52
Db 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 51

RESULT 5
US-10-028-410-3
: Sequence 3, Application US/10028410
: Publication No. US20020160955A1
: GENERAL INFORMATION:
: APPLICANT: Dubaqui, Yves
: APPLICANT: Lowman, Henry
: TITLE OF INVENTION: PROTEIN VARIANTS
: FILE REFERENCE: P1712R1-1
: CURRENT APPLICATION NUMBER: US/10/028-410
: CURRENT FILING DATE: 2001-12-15
: PRIOR APPLICATION NUMBER: US/09/411-04
: NUMBER OF SEQ ID NOS: 6
: SEQ ID NO: 5
: LENGTH: 51
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-028-410-3
Query Match 94.7% Score 279.5 DB 14: Length 51:
Best Local Similarity 94.3% Pred. No. 1.2e-26:
Matches 51: Conservative 0: Mismatches 0: Indels 1: Gaps 1:
Oy 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 52
Db 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 51

RESULT 6
US-09-894-711-18
: Sequence 18, Application US/09894711
: Patent No. US2002017144A1
: GENERAL INFORMATION:
: APPLICANT: Kjelsgaard, Thomas Borlaam
: APPLICANT: Ludvigsen, Svend
: TITLE OF INVENTION: Method for making insulin precursors and
: TITLE OF INVENTION: insulin precursor analogues having improved fermentation
: TITLE OF INVENTION: yield in yeast
: FILE REFERENCE: 6148.400-US
: CURRENT APPLICATION NUMBER: US/09/894-711
: CURRENT FILING DATE: 2001-06-28
: PRIOR APPLICATION NUMBER: PA 2000 00443
: PRIOR FILING DATE: 2000-03-17
: PRIOR APPLICATION NUMBER: PA 1999 01869
: PRIOR FILING DATE: 1999-12-29
: PRIOR APPLICATION NUMBER: 60/211,081
: PRIOR FILING DATE: 2000-06-13
: PRIOR APPLICATION NUMBER: 60/81,450
: PRIOR FILING DATE: 2000-02-10
: PRIOR APPLICATION NUMBER: 09/740,359
: PRIOR FILING DATE: 2000-12-19
: NUMBER OF SEQ ID NOS: 20
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO: 18
: LENGTH: 124
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Synthetic
US-09-894-711-18
Query Match 93.7% Score 275.5 DB 10: Length 124:
Best Local Similarity 94.3% Pred. No. 1.2e-26:
Matches 50: Conservative 1: Mismatches 1: Indels 1: Gaps 1:
```

```
Oy 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 52
Db 72 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 124

RESULT 7
US-09-861-687-19
: Sequence 19, Application US/09861687
: Publication No. US20020193292A1
: GENERAL INFORMATION:
: APPLICANT: Markussen, Jan
: APPLICANT: Jonassen, Ib
: APPLICANT: Havelund, Svend
: APPLICANT: Brandt, Jakob
: APPLICANT: Karlzhals, Peter
: APPLICANT: Hansen, Hertz Per
: APPLICANT: Kaarstoft, Niels Christian
: TITLE OF INVENTION: INSULIN DERIVATIVES
: NUMBER OF SEQUENCES: 26
: PREFERENCE ALGORITHM:
: ADDRESS: No. US20020193292A1 No. US20020193292A1disk of No. US20020193292A1
: STREET: 405 Lexington Avenue, 64th floor
: CITY: New York
: STATE: New York
: COUNTRY: United States of America
: ZIP: 10176 6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/861,687
: FILING DATE: 21-May-2001
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/932,682
: FILING DATE: 16-DEC-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Lambiris, Elias J.
: REGISTRATION NUMBER: 33,728
: REFERENCE/DOCKET NUMBER: 4341.204-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: INFORMATION FOR SEQ ID NO: 19:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 138 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-861-687-19
Query Match 93.7% Score 275.5 DB 10: Length 138:
Best Local Similarity 94.3% Pred. No. 1.4e-26:
Matches 50: Conservative 1: Mismatches 1: Indels 1: Gaps 1:
Oy 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 52
Db 86 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 138

RESULT 8
US-10-066-009A-3
: Sequence 1, Application US/10066009A
: Publication No. US20020165155A1
: GENERAL INFORMATION:
: APPLICANT: Schaffer, Michelle
: APPLICANT: Ullsch, Mark
: APPLICANT: Vajdos, Felix
: TITLE OF INVENTION: CRYSTALLIZATION OF IGF-1
: FILE REFERENCE: P1869R1
```

: CURRENT APPLICATION NUMBER: US/10/066,009A  
 : PRIOR FILING DATE: 2002-06-24  
 : PRIOR APPLICATION NUMBER: US 60/287,072  
 : PRIOR FILING DATE: 2001-04-27  
 : PRIOR APPLICATION NUMBER: US 60/267,977  
 : PRIOR FILING DATE: 2001-02-09  
 : NUMBER OF SEQ ID NOS: 5  
 : SEQ ID NO 3  
 : LENGTH: 50  
 : TYPE: PRT  
 : ORGANISM: Homo sapiens  
 US-10-066-009A-3

Query Match 92.9% Score 273; DB 14; Length 50;  
 Best Local Similarity 96.2% Pred. No. 9,4e-27;  
 Matches 50; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRKRAEDLQGVQLGGPGAGSLQPLALEG 52  
 DB 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRKRAEDLQGVQLGGPGAGSLQPLALEG 50

RESULT 9  
 US-09-878-380-1  
 : Sequence 1; Application US/09878380  
 : Patent No. US20020160435A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Fujirebio Inc.  
 : APPLICANT: KITAJIMA, Sachiko  
 : APPLICANT: KURANO, Yoshihiro  
 : APPLICANT: NAKATSUBO, Kaoru  
 : APPLICANT: NISHIZONO, Isao  
 : TITLE OF INVENTION: Immunoassay For Measuring Human C-Peptide and Kit Therefor  
 : FILE REFERENCE: 0760-0291P  
 : CURRENT APPLICATION NUMBER: US/09/878,380  
 : CURRENT FILING DATE: 2001-06-12  
 : PRIOR APPLICATION NUMBER: JP 2000-174661  
 : PRIOR FILING DATE: 2000-06-12  
 : NUMBER OF SEQ ID NOS: 2  
 : SOFTWARE: Patent In version 3.1  
 : SEQ ID NO 1  
 : LENGTH: 86  
 : TYPE: PRT  
 : ORGANISM: Homo sapiens  
 US-09-878-380-1

Query Match 90.8% Score 263; DB 10; Length 86;  
 Best Local Similarity 60.5% Pred. No. 9,4e-26;  
 Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRKRAEDLQGVQLGGPGAGSLQPLALEG 60  
 DB 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRKRAEDLQGVQLGGPGAGSLQPLALEG 60

QY 31 ----RGIVEQCCTSCISLYOLENYCN 62  
 DB 61 SLQKRGIVEQCCTSCISLYOLENYCN 60

RESULT 10  
 US-09-858-935B-4  
 : Sequence 4; Application US/99858945B  
 : Publication No. US20030069177A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Dubaqui, Yves  
 : APPLICANT: Filvaroff, Ellen  
 : APPLICANT: Lowman, Henry B.  
 : TITLE OF INVENTION: METHOD FOR TREATING CARTILAGE DISORDERS  
 : FILE REFERENCE: P1794R1  
 : CURRENT APPLICATION NUMBER: US/09/858,935B  
 : CURRENT FILING DATE: 2002-07-02  
 : PRIOR APPLICATION NUMBER: US 60/248,985  
 : PRIOR FILING DATE: 2000-11-15

: PRIOR APPLICATION NUMBER: US 60/204,450  
 : PRIOR FILING DATE: 2000-05-16  
 : NUMBER OF SEQ ID NOS: 153  
 : SEQ ID NO 4  
 : LENGTH: 86  
 : TYPE: PRT  
 : ORGANISM: Homo sapiens  
 US-09-858-935B-4

Query Match 90.8% Score 267; DB 11; Length 86;  
 Best Local Similarity 60.5% Pred. No. 9,4e-26;  
 Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRKRAEDLQGVQLGGPGAGSLQPLALEG 60  
 DB 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRKRAEDLQGVQLGGPGAGSLQPLALEG 60

QY 31 ----RGIVEQCCTSCISLYOLENYCN 52  
 DB 61 SLQKRGIVEQCCTSCISLYOLENYCN 86

RESULT 11  
 US-10-028-410-2  
 : Sequence 2; Application US/10028410  
 : Publication No. US20020160955A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Dubaqui, Yves  
 : APPLICANT: Lowman, Henry  
 : TITLE OF INVENTION: PROTEIN VARIANTS  
 : FILE REFERENCE: P1712P1-1  
 : CURRENT APPLICATION NUMBER: US/10/028,410  
 : CURRENT FILING DATE: 2001-12-19  
 : PRIOR APPLICATION NUMBER: US/09/477,924  
 : PRIOR FILING DATE: 2000-01-05  
 : NUMBER OF SEQ ID NOS: 6  
 : SEQ ID NO 2  
 : LENGTH: 86  
 : TYPE: PRT  
 : ORGANISM: Homo sapiens  
 US-10-028-410-2

Query Match 90.8% Score 267; DB 14; Length 86;  
 Best Local Similarity 60.5% Pred. No. 9,4e-26;  
 Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRKRAEDLQGVQLGGPGAGSLQPLALEG 60  
 DB 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRKRAEDLQGVQLGGPGAGSLQPLALEG 60

QY 31 ----RGIVEQCCTSCISLYOLENYCN 52  
 DB 61 SLQKRGIVEQCCTSCISLYOLENYCN 86

RESULT 12  
 US-10-054-473-4  
 : Sequence 4; Application US/10054873  
 : Publication No. US20020164712A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Gan, Zhong RU  
 : TITLE OF INVENTION: Chimeric Protein Containing an Intramolecular Chaperone-Like Sequence  
 : NUMBER OF SEQUENCES: 7  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESS: Townsend and Townsend and Crow LLP  
 : STREET: Two Embarcadero Center, Eighth Floor  
 : CITY: San Francisco  
 : STATE: California  
 : COUNTRY: USA  
 : ZIP: 94111-3834  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/054,873  
FILING DATE: 22-Jan-2002  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/CN98/C0052  
FILING DATE: 31-MAR-1998  
APPLICATION NUMBER: US 09/424,100  
FILING DATE: 11-DEC-2000

ATTORNEY/AGENT INFORMATION:  
NAME: Mycroft, Frank J  
REGISTRATION NUMBER: 46,946  
REFERENCE/DOCKET NUMBER: 020167-00013005

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
LENGTH: 96 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear

MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Escherichia coli  
FEATURE:

SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-10-054-873-4

Query Match 90.8% Score 267.0 Length 96;  
Best Local Similarity 60.5% Pred. No. 1.1e-25;  
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKT- ..... 30

Db 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTR AEDLQGVGVGGPGAGSLOPLALEG 60

QY 31 - - - - -RGIVEQCCTSCISLYOLENYCN 52

Db 61 SLQKRGIVEQCCTSCISLYOLENYCN 86

RESULT 13  
US-09-947-563-4

Sequence 4, Application US/09947563  
Patent No. US2020156214A1  
GENERAL INFORMATION:

APPLICANT: Rubroder, Franz-Josef  
TITLE OF INVENTION: Improved process for obtaining

insulin precursors having correctly bonded cystine bridges

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Finneegan, Hendelison, Farrabow, Garrett &

STREET: 1300 I Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0 Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/947,563

FILING DATE: 07-Sep-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/134,836

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Leslie McDonell

REGISTRATION NUMBER: 34,872

REFERENCE/DOCKET NUMBER: 02481.1600-00000

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 96 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Escherichia coli

FEATURE:

NAME/KEY: Protein

LOCATION: 1-96

SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-09-947-563-4

Query Match 90.8% Score 267.0 Length 96;  
Best Local Similarity 60.5% Pred. No. 1.1e-25;  
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKT- ..... 30

Db 11 FVNOHLCGSHLVEALYLVCGERGFFYTPKTR AEDLQGVGVGGPGAGSLOPLALEG 70

QY 31 - - - - -RGIVEQCCTSCISLYOLENYCN 52

Db 71 SLQKRGIVEQCCTSCISLYOLENYCN 96

RESULT 14

US-09-205-658-125

Sequence 125, Application US/09205658

Patent No. US20010029617A1

GENERAL INFORMATION:

APPLICANT: Ruvkun, Gary

APPLICANT: Oqq, Scott

TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR

FILE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS

FILE REFERENCE: 00786/351004

CURRENT APPLICATION NUMBER: US/09/205,658

CURRENT FILING DATE: 1998-12-03

EARLIER APPLICATION NUMBER: 08/857,076

EARLIER FILING DATE: 1997-05-15

EARLIER APPLICATION NUMBER: 08/889,534

EARLIER FILING DATE: 1997-07-07

EARLIER APPLICATION NUMBER: US98/10080

NUMBER OF SEQ ID NOS: 328

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 125

LENGTH: 110

TYPE: PRT

ORGANISM: Homo sapiens

US-09-205-658-125

Query Match 90.8% Score 267.0 Length 110;  
Best Local Similarity 60.5% Pred. No. 1.2e-25;  
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKT- ..... 30

Db 25 FVNOHLCGSHLVEALYLVCGERGFFYTPKTR AEDLQGVGVGGPGAGSLOPLALEG 84

QY 31 - - - - -RGIVEQCCTSCISLYOLENYCN 52

Db 85 SLQKRGIVEQCCTSCISLYOLENYCN 110

RESULT 15

US-09-815-229-3

Sequence 3, Application US/09815229



```

: Patent No. US20020058614A1
: GENERAL INFORMATION:
: APPLICANT: Filvaroff, Ellen H.
: TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILLAGENOUS DISORDERS
: FILE REFERENCE: P1786R10S
: CURRENT APPLICATION NUMBER: US/09/815,229
: CURRENT FILING DATE: 2001-03-22
: PRIOR APPLICATION NUMBER: US 60/192,103
: PRIOR FILING DATE: 2000-04-24
: NUMBER OF SEQ ID NOS: 17
: SEQ ID NO 3
: LENGTH: 110
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-815,229-3

```

```

Query Match      90.8%   Score 267.   DR 92   Length 110
Best local similarity 60.5%   Posd No. 1 25 25
Matches 52: Conservative 0: Missed 608 04 Indels 54: 048 11
QY      3  FVNQLCGSHLVEALYLVGGERGFFYTPKRI  30
DB      25  FVNQLCGSHLVEALYLVGGERGFFYTPKRI  84
QY      31  --- RGIVEQCCTSIYOLENYN 52
DB      85  SLQKRGIVEQCCTSIYOLENYN 110

```

Search completed: September 15, 2003, 12:25:46  
Job time : 22.8065 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:56:45 ; Search time 8.57348 seconds  
(without alignments)  
883.284 Million cell updates/sec

Title: US-09-423-100-5

Perfect score: 294

Sequence: 1 FVNQHJCGSHLVKALVLCG.....IVFQVCTSGISVLENYCN 52

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Seatched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 283308000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	273.5	93.0	51	1 INWH	insulin - sperm whale
2	273.5	93.0	51	1 INWH	insulin - finback
3	273.5	93.0	51	1 INEL	insulin - elephant
4	274	92.9	96	2 PC7042	epidermal growth f
5	271.5	92.3	51	1 INHY	insulin - hamster
6	268.5	91.3	51	1 NMSSP	insulin - Egyptian
7	267.5	91.0	51	2 A59151	insulin precursor
8	267	90.8	110	1 IPHU	insulin precursor
9	267	90.8	110	2 B42179	insulin precursor
10	267	90.8	110	2 A42174	insulin precursor
11	267	90.8	110	2 JQ0178	insulin precursor
12	263.5	89.6	51	1 INWHIS	insulin - sei whal
13	263.5	89.6	51	1 INGT	insulin - goat
14	263.5	89.6	51	1 INOMA	insulin - Arabian
15	263	89.5	84	1 IPPG	insulin precursor
16	263	89.5	110	1 INRR	insulin precursor
17	262.5	89.3	51	1 INCT	insulin cat
18	262	89.1	110	1 IPDG	insulin precursor
19	261.5	88.9	51	1 INMKSU	insulin - common S
20	260	88.4	110	2 I48166	insulin precursor
21	258.5	87.9	105	1 IPRO	insulin precursor
22	256.5	87.2	51	2 JQ0162	insulin - North Am
23	252.5	85.9	77	1 INSH	insulin precursor
24	252	85.7	86	1 IPHO	insulin precursor
25	251.5	85.5	51	1 INCB	insulin - Chinchil
26	251	85.4	108	2 A39883	insulin precursor
27	250	85.0	108	1 INMS1	insulin 1 precurs
28	249	84.7	110	1 IPRT1	insulin 1 precurs
29	248.5	84.5	51	1 INGS	insulin - goose

#### ALIGNMENTS

##### RESULT 1

INWH

insulin - sperm whale

C:Species: Physeter catodon (sperm whale)

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jul-1999

C:Accession: A93142; A90082

R: Ishihara, Y.; Saito, T.; Ito, T.; Fujino, M.

Nature 181, 1468-1469, 1958

A:Title: Structure of sperm- and sei whale insulins and their breakdown by whale pe

A:Reference number: A93142

A:Accession: A93142

A:Molecule type: protein

A:Residues: 1-30;31-51 <ISH>

R: Harris, J.I.; Sanger, F.; Naughton, M.A.

Arch. Biochem. Biophys. 65, 427-428, 1956

A:Title: Species differences in insulin.

A:Reference number: A90082

A:Accession: A90082

A:Molecule type: protein

A:Residues: 1-30;31-51 <HAR>

C:Superfamily: Insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status experimental <HCH>

F:30-31-51/Product: insulin #status experimental <MAT>

F:31-51/Domain: insulin chain A #status experimental <ACH>

F:7-37;19-50;36-41/Bisulfide bonds: #status predicted

Query Match 93.0% Score 273.5; DB 1; length 51;

Best Local Similarity 96.2%; Pred. No. 1; 5e 24;

Matches 50; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 FVNQHJCGSHLVKALVLCGEGFFYTPKTRGIVGCTSGISVLENYCN 52

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

DB 1 FVNQHJCGSHLVKALVLCGEGFFYTPKTRGIVGCTSGISVLENYCN 51

##### RESULT 2

INWH

insulin - finback whale (tentative sequence)

C:Species: Balaeoptera physalus (finback whale, common forqual)

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jul-1999

C:Accession: A91918

R: Hama, H.; Titani, K.; Sakaki, S.; Narita, K.

J. Biochem. 56, 285-293, 1964

A:Title: The amino acid sequence in fin-whale insulin.

A:Reference number: A91918

A:Accession: A91918

A:Molecule type: protein

A:Residues: 1-30;31-51 <HAM>

C:Superfamily: Insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status experimental <BCH>

A:Residues: 1-30 <MACR>  
A:Accession: A59151



A:Accession: B42179  
A:Molecule type: DNA  
A:Residues: 1-110 <SET>  
A:Cross-references: EMBL:X61092; NID:q22808; PIDN:CAA43405.1; PID:q22809  
A:Note: sequence extracted from NCBI backbone (NCBIN:95185, NCBI:P:95194)  
R:Peterson, J.D.; Nehrllich, S.; Oyer, P.E.; Steiner, D.F.  
J. Biol. Chem. 247, 4866-4871, 1972  
A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog proinsulin  
A:Reference number: A92111; MUID:72258016; PMID:4626369  
A:Accession: A05232  
A:Molecule type: Protein  
A:Residues: 57-87 <PET>  
C:Genetics:  
C:Introns: 63/1  
C:Superfamily: insulin  
C:Keywords: hormone; pancreas  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-54/Domain: insulin chain B #status predicted <BH>  
F:55-87/Domain: insulin connecting peptide #status predicted <CH>  
F:88-110/Domain: insulin chain A #status predicted <AH>  
F:91-96,43-109,95-100/disulfide bonds: #status predicted  
Query Match 90.8%; Score 267; DB 2; Length 110;  
Best Local Similarity 60.5%; Pred. No. 1 6e-23;  
Matches 52; Conservative 0; Mismatches 0; Indels 14; Gaps 1;  
QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKT----- 30  
DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREARDPVGVGLGSGAGSLQPLALEG 84  
QY 31 ----RGIVEQCCTSIQSLYLENYCN 52  
DB 85 SLOKRGIVEQCCTSIQSLYLENYCN 110  
RESULT 13  
Insulin precursor - chimpanzee  
C:Species: Pan troglodytes (chimpanzee)  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 16-Jul-1999  
C:Accession: A42179; S22058  
R:Seino, S.; Bell, G.I.; Li, W.H.  
Mol. Biol. Evol. 9, 193-203, 1992  
A:Title: Sequences of primate insulin genes support the hypothesis of a slower rate of  
A:Reference number: A42179; MUID:9221955; PM: 51560757  
A:Accession: A42179  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-110 <SET>  
A:Cross-references: EMBL:X61089; NID:q38251; PIDN:CAA43403.1; PID:q48252  
A:Note: sequence extracted from NCBI backbone (NCBIN:95067)  
C:Genetics:  
C:Introns: 63/1  
C:Superfamily: insulin  
Query Match 90.8%; Score 267; DB 2; Length 110;  
Best Local Similarity 60.5%; Pred. No. 1 6e-23;  
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;  
QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKT----- 30  
DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREARDPVGVGLGSGAGSLQPLALEG 84  
QY 31 ----RGIVEQCCTSIQSLYLENYCN 52  
DB 85 SLOKRGIVEQCCTSIQSLYLENYCN 110  
RESULT 11  
Insulin precursor - crab-eating macaque  
C:Species: Macaca fascicularis (crab-eating macaque)

C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 16-Jul-1999  
A:Accession: JQ0178  
R:Wetzel, W.; Gronberg, J.; Leinewetter, M.; Wenemayer, F.; Winnacker, E.L.  
Gene 19, 179-183, 1982  
A:Title: The nucleotide sequence of cDNA coding for proinsulin from the primate M.  
A:Reference number: JQ0178; MUID:83080474; PMID:6184262  
A:Accession: JQ0178  
A:Molecule type: mRNA  
A:Residues: 1-110 <SET>  
A:Cross-references: GB:J00336; NID:q342121; PIDN:AAA36849.1; PID:q342122  
C:Superfamily: insulin  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-54/Domain: insulin chain B #status predicted <BH>  
F:55-87/Domain: insulin connecting peptide #status predicted <CH>  
F:88-110/Domain: insulin chain A #status predicted <AH>  
F:91-96,43-109,95-100/disulfide bonds: #status predicted  
Query Match 90.8%; Score 267; DB 2; Length 110;  
Best Local Similarity 60.5%; Pred. No. 1 6e-23;  
Matches 52; Conservative 0; Mismatches 0; Indels 14; Gaps 1;  
QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKT----- 30  
DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREARDPVGVGLGSGAGSLQPLALEG 84  
QY 31 ----RGIVEQCCTSIQSLYLENYCN 52  
DB 85 SLOKRGIVEQCCTSIQSLYLENYCN 110  
RESULT 12  
Insulin - sei whale  
C:Species: Balenoptera borealis (sei whale)  
C:Date: 13-Jul-1981 #sequence\_revision 13-Jul-1981 #text\_change 16-Jul-1999  
C:Accession: A01582  
R:Ushihara, Y.; Saito, T.; Ito, Y.; Fujino, M.  
Nature 181, 1468-1469, 1958  
A:Title: Structure of sperm- and sei-whale insulins and their breakdown by whale pep.  
A:Reference number: A93142  
A:Accession: A01582  
A:Molecule type: protein  
A:Residues: 1-30,31-51 <ISH>  
C:Superfamily: insulin  
C:Keywords: hormone; pancreas  
F:1-30/Domain: insulin chain B #status experimental <BH>  
F:31-30,31-51/Product: insulin #status experimental <MAT>  
F:41-51/Domain: insulin chain A #status experimental <AH>  
F:7-37,19-50,36-41/disulfide bonds: #status predicted  
Query Match 89.6%; Score 263.5; DB 1; Length 51;  
Best Local Similarity 92.3%; Pred. No. 2 1e-24;  
Matches 48; Conservative 0; Mismatches 3; Indels 1; Gaps 1;  
QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREARDPVGVGLGSGAGSLQPLALEG 52  
DB 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRREARDPVGVGLGSGAGSLQPLALEG 51  
RESULT 13  
Insulin - goat  
C:Species: Capra aegagrus hircus (domestic goat)  
C:Date: 13-Jul-1981 #sequence\_revision 13-Jul-1981 #text\_change 16-Jul-1999  
C:Accession: A01586  
R:Smith, I.F.  
Am. J. Med. 40, 662-666, 1966  
A:Title: Species variation in the amino acid sequence of insulin.  
A:Reference number: A90029; MUID:66160119; PMID:5949593  
A:Accession: A01586  
A:Molecule type: protein  
A:Residues: 1-30,31-51 <SMI>

C:Superfamily: insulin  
 C:Keywords: hormone; pancreas  
 F:1-30/Domain: insulin chain B #status experimental <BCH>  
 F:1-30/1-51/Product: insulin #status experimental <MAT>  
 F:31-51/Domain: insulin chain A #status experimental <ACH>  
 F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 89.6%; Score 263.5; DB 1; Length 51;  
 Best local Similarity 90.4%; Pred. No. 2,1e-23;  
 Matches 47; Conservative 1; Mismatches 3; Indels 1; Gaps 1;  
 Oy 1 FVNOHLGGSHLVEALYLVGGERGFFYTPKTR:IVEQCCTSIICSLYLENYCN 52  
 I:||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 1 FVNOHLGGSHLVEALYLVGGERGFFYTPKA:GIVEQCAGVCSLYOLENYCN 51

## RESULT 14

INCHA Arabian camel (Camelus dromedarius)  
 C:Species: Camelus dromedarius (Arabian camel)  
 C:Date: 31-Mar-1992 #sequence\_revision 11 Mar 92 atoxl change 16 Jul 99  
 C:Accession: A92782  
 R:Danho, W.O.  
 J. Fac. Med. Riyadh 14, 16-28, 1972  
 A:Title: The isolation and characterization of insulin of camel (Camelus dromedarius).  
 A:Reference number: A92782

A:Accession: A92782  
 A:Molecule type: protein  
 A:Residues: 1-30;31-51 <DAN>  
 C:Superfamily: insulin  
 C:Keywords: hormone; pancreas  
 F:1-30/Domain: insulin chain B #status experimental <BCH>  
 F:1-30;31-51/Product: insulin #status experimental <MAT>  
 F:31-51/Domain: insulin chain A #status experimental <ACH>  
 F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 89.6%; Score 263.5; DB 1; Length 51;  
 Best local Similarity 90.4%; Pred. No. 2,1e-23;  
 Matches 47; Conservative 1; Mismatches 3; Indels 1; Gaps 1;  
 Oy 1 FVNOHLGGSHLVEALYLVGGERGFFYTPKTR:IVEQCCTSIICSLYLENYCN 52  
 I:||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 1 FVNOHLGGSHLVEALYLVGGERGFFYTPKA:GIVEQCAGVCSLYOLENYCN 51

## RESULT 15

IPPG Insulin precursor - pig  
 C:Species: Sus scrofa domestica (domestic pig)  
 C:Date: 22-Jun-1991 #sequence\_revision 22-Jun 91 #text\_change 16 Jul 1999  
 C:Accession: A01583; A94572; S16472; A60845; 0845  
 R:Chanoe, R.E.; Ellis, R.M.; Bruner, W.W.  
 Science 161, 165-167, 1968  
 A:Title: Porcine proinsulin: characterization and amino acid sequence.  
 A:Reference number: A94240; M010:68286445; pp 13657063

A:Accession: A01583  
 A:Molecule type: protein  
 A:Residues: 1-34;35-84 <CHA>  
 R:Chanoe, R.E.  
 Submitted to the Atlas, July 1970  
 A:Reference number: A94572  
 A:Accession: A94572  
 A:Molecule type: protein  
 A:Residues: 1-84 <CH2>  
 R:Brown, H.; Sanger, F.; Kitai, R.  
 Biochem. J. 60, 566-565, 1955  
 A:Title: The structure of pig and sheep insulin.  
 A:Reference number: A90344  
 A:Accession: S16492  
 A:Molecule type: protein  
 A:Residues: 1-30;31-51 <BRO>  
 R:Snell, L.; Damgaard, U.  
 Horm. Metab. Res. 20, 476-480, 1988

A:Title: Proinsulin heterogeneity in pigs.  
 A:Reference number: A60835; M010:69032178; PMID:3181865  
 A:Accession: A60835  
 A:Molecule type: protein  
 A:Residues: 33-38,40-62 <SNE>  
 A:Note: the authors report the characterization of a connecting peptide variant lack  
 A:Accession: A60835  
 A:Molecule type: protein  
 A:Residues: 33-62 <SN2>  
 R:Blundell, T.; Dodson, G.; Rodkin, D.; Mercola, D.  
 Adv. Protein Chem. 26, 279-402, 1972  
 A:Title: Insulin: the structure in the crystal and its reflection in chemistry and i  
 A:Reference number: A90017  
 A:Contents: annotation: X-ray crystallography, 1.9 angstroms  
 C:Superfamily: insulin  
 C:Keywords: hormone; pancreas  
 F:1-30/Domain: insulin chain B #status experimental <BCH>  
 F:1-30;64-84/Product: insulin #status experimental <MAT>  
 F:33-64/Domain: connecting peptide #status experimental <CPEP>  
 F:64-84/Domain: insulin chain A #status experimental <ACH>  
 F:7-70,19-41,69-74/Disulfide bonds: #status experimental

Query Match 89.5%; Score 263; DB 1; Length 84;  
 Best local Similarity 60.7%; Pred. No. 3.6e-23;  
 Matches 51; Conservative 0; Mismatches 1; Indels 32; Gaps 1;  
 Oy 1 FVNOHLGGSHLVEALYLVGGERGFFYTPKTR:IVEQCCTSIICSLYLENYCN 52  
 I:||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 1 FVNOHLGGSHLVEALYLVGGERGFFYTPKRA:ENPOAGAVEIGGIGGLQALALEGPP 60

Oy 31 --RGIVPQCCTSIICSLYLENYCN 52  
 I:||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 61 QKRGIVPQCCTSIICSLYLENYCN 84

Search completed: September 15, 2003, 12:04:20  
 Job time : 9.57348 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:54:00 ; Search time 4.84598 seconds  
(without alignments)  
504.633 Million cell updates/sec

Title: US-09-423-100-5

Perfect score: 294

Sequence: 1 FVNQHICGSHLVEALYLVCGR.....IVEGCTSTCSLYOLENYCN 52

Scoring table: BLOSUM62

Gapof: 10.0 , Gapext: 0.5

Searched: 12786.4 seqs, 479,263,95 residues

Total number of hits satisfying chosen parameters: 12786.4

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	273.5	93.0	51	INS_RALPH	P01312 balaenopter
2	273.5	93.0	51	INS_ELMA	P01316 elephas max
3	268.6	91.3	51	INS_ACOCA	P03324 acorys cabi
4	267	90.8	110	INS_CERAE	P04047 ceropithec
5	267	90.8	110	INS_HUMAN	P01308 homo sapien
6	267	90.8	110	INS_MACEA	P30406 macaca fasc
7	267	90.8	110	INS_PANTR	P30410 pan troglod
8	263.5	89.6	51	INS_RALHO	P01314 balaenopter
9	263.5	89.6	51	INS_CANDR	P01320 camelus dro
10	264.5	89.6	51	INS_CAPHI	P01319 capra hircu
11	263	89.5	108	INS_FIG	P01315 sus scrofa
12	263	89.5	110	INS_RAHIT	P01311 oryctolagus
13	263	89.5	110	INS_SPTTR	G91X13 spermophilu
14	262.5	89.3	51	INS_FELCA	P06306 felis silve
15	262	88.4	110	INS_CANFA	P01321 canis fami
16	260	88.4	110	INS_CRILO	P01313 cricetus
17	258.5	87.9	105	INS_BOVIN	P01317 bos taurus
18	257	87.4	110	INS_PCAOR	G62587 psammomys c
19	256.5	87.2	51	INS_DUMA	P01309 didelphis m
20	254.5	86.6	105	INS_SHEEP	P01318 ovis aries
21	252	85.7	86	INS_HORSE	P01310 equus cabal
22	251.5	85.5	51	INS_CHIRR	P01327 chinchilla
23	251	85.4	108	INS_AOTTR	P10604 aotus trivi
24	250	85.0	108	INS_MOUSE	P01325 mus musculu
25	249	84.7	110	INS1_RAT	P01322 rattus norv
26	248.5	84.5	51	INS_ANGAN	P07454 anser anser
27	248	84.4	110	INS2_MOUSE	P01326 mus musculu
28	248	84.4	110	INS2_RAT	P01323 rattus norv
29	246	83.7	52	INS_ACIGU	P81423 acipeenser g
30	244.5	83.2	51	INS_HYSCR	P01328 hystrix cri
31	244.5	83.2	51	INS1_TRASC	P31887 trachemys s
32	238.5	81.1	103	INS_SELRF	P51463 selasphorus
33	235.5	80.1	51	INS_ORNAN	O91qy7 ornithorhyn

RESULT 1

INS_RALPH	INS_RALPH	STANDARD	PK1	5, AA	
AC	P01312				
DI	21-JUL-1986 (Ref. 01, created)				
DT	21-JUL-1986 (Ref. 01, last sequence update)				
DI	01-OCT-1996 (Ref. 34, last annotation update)				
DE	Insulin.				
GN	INS.				
OS	Balaenoptera physalus (Finback whale) (Common rorqual), and				P01332 gallus gall
OS	Physeter catodon (Sperm whale) (Physeter macrocephalus).				P01333 anas platyr
OC	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:				P12703 alligator m
OC	Mammalia: Eutheria: Cetartiodactyla: Cetacea: Mysticeti:				P09476 icpistosteus
CC	Balaenopteridae: Balaenoptera.				P12708 zaocys dhum
OX	NCBI_TaxID=9770, 9755;				P01334 crotalus at
RP	[1]				P12706 xenopus lae
RP	PARTIAL SEQUENCE.				P12707 xenopus lae
RC	SPECIES-B.physalus;				P14806 petromyzon
RA	Hama H., Titani K., Sakaki S., Narita K.;				P09477 platichthys
RT	"The amino acid sequence in fin-whale insulin.";				P29335 amia calva
KL	J. Biochem. 56:285-293(1964).				P21187 oncorhynchus
RN	[2]				
RP	SPECIES-P.catodon;				
RC	SPECIES-P.catodon;				
RA	Ishihara Y., Saito T., Ito Y., Fujino M.;				
RT	"Structure of sperm- and sei-whale insulins and their breakdown by				
RL	whale pepsin.";				
RN	Nature 181:1468-1469(1958).				
RP	[3]				
RC	SPECIES-P.catodon;				
RA	Harris J. I., Sanger F., Naughton M. A.;				
RT	"Species differences in insulin.";				
CC	Arch. Biochem. Biophys. 65:427-448(1956).				
CC	FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT				
CC	INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND				
CC	FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE				
CC	CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.				
CC	FUNCTION: HETEROLOGOUS OF A B CHAIN AND AN A CHAIN LINKED BY TWO				
CC	SULFIDE BONDS.				
CC	FUNCTION: SUBCELLULAR LOCATION: Secreted.				
CC	FUNCTION: SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.				
DR	PIR: A91918; IWHNP.				
DR	PIR: A91142; IWHNP.				
DR	RSSP: P01317; IAPH.				
DR	InterPro: IPR024825; Ins/IGF/relax.				
DR	SMART: SM00078; IIGF. 1.				
DR	PROSITE: PS00262; INSULIN; 1				
KW	Insulin family; Hormone; Glucose metabolite.				
FT	CHAIN 1 30				
FT	NON-CONS 30 31				
FT	CHAIN 30 31				
FT	DISULFID 7 37				
FT	DISULFID 19 50				
FT	DISULFID 36 41				
FT	SEQUENCE 51 AA; 5766 MW; 9007B514691A7CID0_CRC64;				

```

Query Match      93.0%  Score 273.5;  DB 1;  Length 51;
Best Local Similarity 96.2%  Pred. No. 4.9e-27;
Matches 50;  Conservative 0;  Mismatches 1;  Gaps 1;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCSLYLENYCN 52
    |||||||
DB 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCSLYLENYCN 51

RESULT 2
INS_ELEMA
ID INS_ELEMA STANDARD; PRT: 51 AA.
AC P01316;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin
GN INS.
OS Elephas maximus (Indian elephant).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Proboscidea; Elephantidae; Elephas.
OX NCBI_TaxID=9783;
RN [1]
RP SEQUENCE.
RX MEDLINE=66160119; PubMed=5949593;
RA Smith L.F.;
RT "Species variation in the amino acid sequence of insulin."
RL Am. J. Med. 40:662-666(1966).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR InterPro: IPR004825; Ins/IGF/relax.
DR SMART: SM00078; IIGF.
DR PROSITE: PS00262; INSULIN.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 32 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41 INTERCHAIN.
SQ SEQUENCE 51 AA: 5752 MW: 500785.308457060 CRC64;

Query Match      93.0%  Score 273.5;  DB 1;  Length 51;
Best Local Similarity 96.2%  Pred. No. 4.9e-27;
Matches 49;  Conservative 1;  Mismatches 1;  Gaps 1;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCSLYLENYCN 52
    |||||||
DB 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCSLYLENYCN 51

RESULT 1
INS_ACOXA
ID INS_ACOXA STANDARD; PRT: 51 AA.
AC P01324;
DT 21 JUL 1986 (Rel. 01, Created)
DT 21 JUL 1986 (Rel. 01, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Insulin.
GN INS.
OS Acomys cahirinus (Egyptian spiny mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Acomys.

```

```

OX NCBI_TaxID=10068;
RN [1]
RP COMPOSITION.
RX MEDLINE=72189454; PubMed=5028210;
RA Buenzli H.F., Hummel R.E.;
RT "Isolation and partial structural analysis of insulin from mouse (Mus musculus) and spiny mouse (Acomys cahirinus).".
RL Hoppe-Seyler's Z. Physiol. Chem. 353:444-450(1972).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR InterPro: IPR004825; Ins/IGF/relax.
DR SMART: SM00078; IIGF.
DR PROSITE: PS00262; INSULIN.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 32 INTERCHAIN (BY SIMILARITY).
FT DISULFID 19 50 INTERCHAIN (BY SIMILARITY).
FT DISULFID 36 41 BY SIMILARITY.
SQ SEQUENCE 51 AA: 5768 MW: 992808629647040 CRC64;

Query Match      91.3%  Score 268.5;  DB 1;  Length 51;
Best Local Similarity 92.1%  Pred. No. 26-26;
Matches 48;  Conservative 3;  Mismatches 0;  Indels 1;  Gaps 1;

OY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCSLYLENYCN 52
    |||||||
DB 1 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCSLYLENYCN 51

RESULT 4
INS_CERAE
ID INS_CERAE STANDARD; PRT: 110 AA.
AC P30407; P01309;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin precursor.
GN INS.
OS Cercopithecus aethiops (green monkey) (Primate).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=9534;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92219953; PubMed=1560757;
RA Seino S., Rell G.L., Li W.;
RT "Sequences of primate insulin genes support the hypothesis of a
RT slower rate of molecular evolution in humans and apes than in
RT monkeys."
RL Mol. Biol. Evol. 9:194-203(1992).
RN [2]
RP SEQUENCE OF 57-87.
RX MEDLINE=72258016; PubMed=4626369;
RA Peterson J.D., Nehrlich S., Oyer P.E., Steiner D.F.;
RT "Determination of the amino acid sequence of the monkey, sheep, and
RT dog proinsulin C-peptides by a semi-micro Edman degradation
RT procedure."
RL J. Biol. Chem. 247:4866-4871(1972).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.

```



```

CC      -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC      DISULFIDE BONDS.
CC      -1- SURCELLULAR LOCATION: Secreted.
CC      -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL: X61092; CAA43405.1; .
CC      DR      PIR: R42179; R42179.
CC      DR      HSSP: P01308; I410.
CC      DR      InterPro: IPR004825; Ins/IGF/relax.
CC      DR      Pfam: PF00479; Ins_H1007.
CC      DR      SMART: SM00498; IUGR1.
CC      DR      PROSITE: PS00262; INSULIN 1.
CC      KW      Insulin family; Hormone; Glucose metabolism; Signal.
CC      FT      SIGNAL: 1 24
CC      FT      CHAIN: 25 54 INSULIN B CHAIN.
CC      FT      PROPEP: 57 87 C PEPTIDE.
CC      FT      CHAIN: 90 110 INSULIN A CHAIN.
CC      FT      DISULFID: 31 95 INTERCHAIN.
CC      FT      DISULFID: 43 105 INTERCHAIN.
CC      FT      DISULFID: 95 100
CC      SQ      SEQUENCE 110 AA; 95AIF54BF7B247F9 CRC64;
CC
CC      Query Match 90.8%; Score 267; DP 1; Length 110;
CC      Best Local Similarity 60.5%; Pred. No. 6 30-26;
CC      Matches 52; Conservative 0; Mismatches 6; Indels 14; Gaps 1;
CC
QY      1 FYNQHLGSHLVEAIVLCGERGFFYTPKTI 30
DB      25 FYNQHLGSHLVEAIVLCGERGFFYTPKTI KEADHVGQVGVGGGAGSLQPLAEG 84
QY      31 -RGIVEQCCTSCISLYOLENYCN 52
DB      85 SLQKRGIVEQCCTSCISLYOLENYCN 110
CC
RESULT 5
INS_HUMAN
ID      INS_HUMAN STANDARD: PR3; 11; AA.
AC      P01308;
DT      21-JUL-1986 (Ref. 01, Created)
DT      21-JUL-1986 (Ref. 01, Last sequence update)
DT      15-SEP-2003 (Ref. 42, Last annotation update)
DE      Insulin precursor.
GN      INS.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID:9606;
RN      [1]
RX      MEDLINE-80120725; PubMed-6243748;
RA      Bell G.I., Pictet R.L., Rutter W.J., Cordell B., Tischer E.,
RA      Goodman H.M.;
RT      "Sequence of the human insulin gene.";
RL      Nature 284:26-32(1980).
RN      [2]
RX      MEDLINE-80236313; PubMed-6248962;
RA      Ullrich A., Dull T.J., Gray A., Brosius J., Sures I.;
RT      "Genetic variation in the human insulin gene.";
RL      Science 209:612-615(1980).
RN      [3]
RX      MEDLINE-80054779; PubMed-503234;
RA      Bell G.I., Swain W.F., Pictet R.L., Cordell B., Goodman H.M.,

```

```

RA      Rutter W.J.;
RT      "Nucleotide sequence of a cDNA clone encoding human preproinsulin.";
RL      Nature 282:525-527(1979).
RN      [4]
RX      MEDLINE-80147417; PubMed-6927840;
RA      Sures I., Goeddel D.V., Gray A., Ullrich A.;
RT      "Nucleotide sequence of human preproinsulin complementary DNA.";
RL      Science 208:57-59(1980).
RN      [5]
RX      MEDLINE-93364424; PubMed-6458440;
RA      Lucassen A.M., Reil J.I., Julier C., Rathrop M.;
RT      "Susceptibility to insulin dependent diabetes mellitus maps to a 4.1
RT      kb segment of DNA spanning the insulin gene and associated VNTR.";
RL      Nat. Genet. 4:305-310(1993).
RN      [6]
RX      MEDLINE-22348252; PubMed-12477342;
RA      Sittasberg K.L., Frittsold E.A., Goto S., Loefer J.G.,
RA      Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.B.,
RA      Altschul S.F., Zuberq B., Buetow K.H., Schaffer C.F., Bhat N.K.,
RA      Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA      Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA      Stapleton M., Soares M.B., Ronald M.F., Casavant T.L., Scheetz T.E.,
RA      Brownstein M.J., Udlin T.B., Toshiyuki S., Carninci P., Prange C.,
RA      Raha S.S., Loughran N.A., Peters G., Abramson R.D., Mullaly S.J.,
RA      Hosak S.A., McEwan P.J., McKernan K., Malek J.A., Gunaratne P.H.,
RA      Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
RA      Villalón D.K., Mizny D.M., Sodergren E.J., Li X., Gibbs R.A.,
RA      Fahy J., Helton E., Kettelman M., Mauden A., Rodrigues S., Sanchez A.,
RA      Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA      Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA      Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA      Butterfield Y.S.N., Krzywinski M.L., Skalska U., Small D.E.,
RA      Schurch A., Schein J.E., Jones S.J.M., Maffa M.A.;
RT      "Generation and initial analysis of more than 15,000 full-length
RT      human and mouse cDNA sequences.";
RL      Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN      [7]
RX      MEDLINE-71116410; PubMed-5101771;
RA      Oyer P.E., Cho S., Peterson J.D., Steiner D.F.;
RT      "Studies on human proinsulin. Isolation and amino acid sequence of
RT      the human pancreatic C-peptide.";
RL      J. Biol. Chem. 246:1375-1386(1971).
RN      [10]
RX      MEDLINE-71257722; PubMed-5560404;
RA      KO A., Smyth D.G., Markussen J., Sundby F.;
RT      "The amino acid sequence of the C-peptide of human proinsulin.";
RL      Eur. J. Biochem. 26:190-199(1971).
RN      [11]
RX      MEDLINE-75077277; PubMed-4443293;
RA      Sieber P., Kamber B., Hartmann A., Joehl A., Riniker B., Rittel W.;
RT      "Total synthesis of human insulin under directed formation of the
RT      disulfide bonds.";
RL      Helv. Chim. Acta 57:2617-2621(1974).
RN      [12]

```

RP SYNTHESIS OF 57-87.  
 RX MEDLINE-75940007; PubMed-4803504;  
 RA Naithani V.K.;  
 RT "Studies on polypeptides, IV. The synthesis of C-peptide of human  
 RT proinsulin";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 354:659-672(1973).  
 RN [13]  
 RP SYNTHESIS OF 65-69 AND 70-73.  
 RX MEDLINE-73161263; PubMed-4698555;  
 RA Geiger R., Volk A.;  
 RT "Synthesis of peptides with the properties of human proinsulin C-  
 RT peptides (hc peptide). 3. Synthesis of the sequences 14-17 and 9-13  
 RT of human proinsulin C-peptides";  
 RL Chem. Ber. 106:199-205(1973).  
 RN [14]  
 RP SYNTHESIS OF 84-87.  
 RX MEDLINE-73161261; PubMed-4698555;  
 RA Geiger R., Jaeger G., Koenig W., Tschopp C.;  
 RT "Synthesis of peptides with the properties of human proinsulin C-  
 RT peptides (hc peptide). 1. Scheme for the synthesis and preparation of  
 RT the sequence 28-31 of human proinsulin C-peptide";  
 RL Chem. Ber. 106:188-192(1973).  
 RN [15]  
 RP VARIANT LOS ANGELES SER-48.  
 RX MEDLINE-84016053; PubMed-6312455;  
 RA Haneda M., Chan S.J., Kwok S.C.M., Kubota  
 RT "Studies on mutant human insulin genes: identification and sequence  
 RT analysis of a gene encoding [SerB24]insulin";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:6356-6358(1983).  
 RN [16]  
 RP VARIANT LOS ANGELES SER-48 AND CHICAGO 10-49.  
 RX MEDLINE-84170233; PubMed-6424113;  
 RA Shoelson S., Flicker M., Haneda M., Nahay A., Musso G., Kaiser E.L.,  
 RA Rubenstein A.H., Tager H.;  
 RT "Identification of a mutant human insulin predicted to contain a  
 RT serine for-phenylalanine substitution";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:7390-7394(1983).  
 RN [17]  
 RP VARIANT PROVIDENCE ASP-34.  
 RX MEDLINE 87175640; PubMed-3470784;  
 RA Chan S.J., Seino S., Gruppese P.A., Schwartz R., Steiner D.F.;  
 RT "A mutation in the B chain coding region is associated with impaired  
 RT proinsulin conversion in a family with hyperproinsulinemia";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:2194-2197(1987).  
 RN [18]  
 RP VARIANT WAKAYAMA LEU-92.  
 RX MEDLINE-87058122; PubMed-3547011;  
 RA Sakura H., Iwamoto Y., Sakamoto Y., Kuzuya T., Hirata H.;  
 RT "Structurally abnormal insulin in a diabetic patient. Characterization  
 RT of the mutant insulin A3 (Val-->Leu) isolated from the pancreas";  
 RL J. Clin. Invest. 78:1666-1672(1986).  
 RN [19]  
 RP VARIANT HIS 89.  
 RX MEDLINE-90317021; PubMed-2196279;  
 RA Barbetti F., Raben N., Kadwaki T., Rama A., Accilli G., Gabbay K.H.,  
 RA Weresch J.A., Taylor S.L., Roth J.;  
 RT "Two unrelated patients with familial hyperproinsulinemia due to a  
 RT mutation substituting histidine for arginine at position 65 in the  
 RT proinsulin molecule: identification of the mutation by direct  
 RT sequencing of genomic deoxyribonucleic acid amplified by polymerase  
 RT chain reaction";  
 RL J. Clin. Endocrinol. Metab. 71:164-169(1990).  
 RN [20]  
 RP VARIANT HIS 89.  
 RX MEDLINE-85261996; PubMed-4019786;  
 RA Shibasaki Y., Kawakami T., Kanazawa Y., Akauma Y., Takaku F.;  
 RT "Posttranslational cleavage of proinsulin is blocked by a point  
 RT mutation in familial hyperproinsulinemia";  
 RL J. Clin. Invest. 76:378-380(1985).  
 RN [21]  
 RP VARIANT KYOTO LEU-89.  
 RX MEDLINE-92291307; PubMed-1601997;  
 RA Yano H., Kilano N., Morimoto M., Polonsky K.S., Imura H., Seino Y.;

\*A novel point mutation in the human insulin gene giving rise to  
 RT hyperproinsulinemia (proinsulin Kyoto);  
 RL J. Clin. Invest. 89:1902-1907(1992).  
 RN [22]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE-91104966; PubMed-2271664;  
 RA Hua Q.-X., Weiss M.A.;  
 RT "Toward the solution structure of human insulin: sequential 2D 1H NMR  
 RT assignment of a des-pentapeptide analogue and comparison with crystal  
 RT structure";  
 RL Biochemistry 29:10545-10555(1990).  
 RN [23]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE-91242467; PubMed-2046420;  
 RA Hua Q.-X., Weiss M.A.;  
 RT "Comparative 2D-NMR studies of human insulin and des-pentapeptide  
 RT insulin: sequential resonance assignment and implications for protein  
 RT dynamics and receptor recognition";  
 RL Biochemistry 30:5536-5545(1991).  
 RN [24]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE 91265527; PubMed-1646635;  
 RA Hua Q.-X., Weiss M.A.;  
 RT "Two-dimensional NMR studies of Des-(B26-B30) insulin: sequence-  
 RT specific resonance assignments and effects of solvent composition";  
 RL Biochim. Biophys. Acta 1078:101-110(1991).  
 RN [25]  
 Query Match 90.89; Score 267; DB 1; Length 110;  
 Best local Similarity 60.58; Pred. No. 6, 30-26;  
 Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;  
 QY 1 FVNHLCGSHLVYALYVNGERGFYTPK? ..... 30  
 DB 25 FVNHLCGSHLVYALYVNGERGFYTPKRRAREGLQVQVELGGGAGSLQFLALRG 84  
 QY 31 ----RGIVEGCTGCSLSLYOLENYCN 52  
 DB 85 SLQKRGIVEGCTGCSLSLYOLENYCN 110  
 RESULT: 6  
 INS\_MACFA  
 ID INS\_MACFA STANDARD; PRT: 110 AA.  
 AC P30406; P01309;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DI 13-AUG-1987 (Rel. 05, Last sequence update)  
 DE 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Insulin precursor.  
 GN INS.  
 OS Macaca fascicularis (Grab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecoidea; Macaca.  
 OX NCBI TaxID:9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-83080474; PubMed-6184262;  
 RA Wetckam W., Gronenberg J., Leineweber M., Wengrenmayer F.;  
 RA Winnacker E.-L.;  
 RT "The nucleotide sequence of cDNA coding for preproinsulin from the  
 RT primate Macaca fascicularis";  
 RL Gene 19:179-183(1982).  
 RN [2]  
 CC FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: J00336; AAA36849.1; .  
 DR J00178; J00178.  
 DR HSP: P01308; IAI0.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR PRINTS: PR00277; INSULIN.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54 INSULIN B CHAIN.  
 FT PROPEP 57 87 C PEPTIDE.  
 FT CHAIN 90 110 INSULIN A CHAIN.  
 FT DISULFID 31 96 INTERCHAIN.  
 FT DISULFID 43 109 INTERCHAIN.  
 FT DISULFID 95 100  
 SQ SEQUENCE 110 AA: 11991 MW: 836E33A80A420F9 CRC64;

Query Match 90.8% Score 267; DB 1; Length 110;  
 Best Local Similarity 60.5%; Pred. No. 6,3e-26;  
 Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 1 FVNOHLGCSHLVEALVLCGERGFYTKT ----- 30

DB 25 FVNOHLGCSHLVEALVLCGERGFYTKT----- 84

QY 31 ---RGIVQGCCTSCSYLYENYCN 52

DB 85 SLQKRGIVQGCCTSCSYLYENYCN 110

#### RESULT 7

INS\_PANTR  
 ID INS\_PANTR STANDARD; PRT: 110 AA.  
 AC P30410;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Insulin precursor.  
 GN INS.  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 OX NCBI\_TaxID=9598;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE:92214453; PubMed:1560757;  
 RA Seino S., Bell G.I., Li W.;  
 RT "Sequences of primate insulin genes support the hypothesis of a slower rate of molecular evolution in humans and apes than in monkeys".  
 RL Mol Biol Evol. 9:193-203(1992).  
 CC !- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC !- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.  
 CC !- SUBCELLULAR LOCATION: Secreted.  
 CC !- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: X61089; CAA43403.1; .  
 DR PIR: A42179; A42179.  
 DR PDB: 1EFE; 29-MAR-00.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR PRINTS: PR00277; INSULIN.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.

KW Insulin family; Hormone; Glucose metabolism; Signal; 3D-structure.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54 INSULIN B CHAIN.  
 FT PROPEP 57 87 C PEPTIDE.  
 FT CHAIN 90 110 INSULIN A CHAIN.  
 FT DISULFID 31 96 INTERCHAIN.  
 FT DISULFID 43 109 INTERCHAIN.  
 FT DISULFID 95 100  
 SQ SEQUENCE 110 AA: 12025 MW: 41E88E7963DEFF5 CRC64;

Query Match 90.8% Score 267; DB 1; Length 110;  
 Best Local Similarity 60.5%; Pred. No. 6,3e-26;  
 Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 1 FVNOHLGCSHLVEALVLCGERGFYTKT ----- 30

DB 25 FVNOHLGCSHLVEALVLCGERGFYTKT----- 84

QY 31 ---RGIVQGCCTSCSYLYENYCN 52

DB 85 SLQKRGIVQGCCTSCSYLYENYCN 110

#### RESULT 8

INS\_BALRO  
 ID INS\_BALRO STANDARD; PRT: 51 AA.  
 AC P01314;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Insulin.  
 GN INS.  
 OS Balanoptera borealis (Sei whale).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;  
 OC Balanopteridae; Balanoptera.  
 OX NCBI\_TaxID=9768;  
 RN [1]  
 RP SEQUENCE.  
 RA Ishihara Y., Saito T., Ito Y., Fujino M.;  
 RT "Structure of sperm- and sei-whale insulins and their breakdown by whale pepsin".  
 RL Nature 181:1468-1469(1958).  
 CC !- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC !- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.  
 CC !- SUBCELLULAR LOCATION: Secreted.  
 CC !- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC PIR: A01582; INWHIS.  
 DR HSP: P01317; IAPH.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism.  
 FT CHAIN 1 30 INSULIN B CHAIN.  
 FT NON\_CONS 30 31  
 FT CHAIN 31 51 INSULIN A CHAIN.  
 FT DISULFID 7 37 INTERCHAIN.  
 FT DISULFID 19 50  
 FT DISULFID 36 41



RP REVISION TO 59.  
 RA Chance R.E.;  
 RL Submitted (JUL-1970) to the PIR data bank.  
 [5]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RA Blundell T.L., Dodson G.G., Hodgkin D., Mercola D.;  
 RT Insulin. The structure in the crystal and its reflection in  
 chemistry and biology.;  
 RL Adv. Protein Chem. 26:279-402(1972).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).  
 RA Isaacs N.W., Agarwal R.C.;  
 RT Experience with fast Fourier least squares in the refinement of the  
 crystal structure of rhombohedral 2-zinc insulin at 1.5-A  
 resolution.;  
 RL Acta Crystallogr. A 34:782-791(1978).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).  
 RX MEDLINE 8909418; PubMed 2465486;  
 RA Baker E.N., Blundell T.L., Cutcliffe J.P., Cutcliffe S.M., Dodson G.G.;  
 RA Dodson G.G., Crowfoot Hodgkin D.M., Suck D., Isaacs N.W.;  
 RA Reynolds C.D., Sakabe K., Sakabe N., Vijayan N.M.;  
 RT The structure of 2Zn pig insulin crystals at 1.5 A resolution.;  
 RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 319:369-456(1988).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RX MEDLINE 92126280; PubMed 1772633;  
 RA Balchmidt P., Hansen F.B., Dodson E., Dodson G., Korbey F.;  
 RT Structure of porcine insulin co-crystallized with glupeine Z.;  
 RL Acta Crystallogr. B 47:975-986(1991).  
 RN [9]  
 RP X-RAY CRYSTALLOGRAPHY.  
 RX MEDLINE 91222450; PubMed 2025410;  
 RA Badger J., Harris M.R., Reynolds G.D., Evans A.C., Dodson E.;  
 RA Dodson G.G., North A.C.T.;  
 RT Structure of the pig insulin dimer in the cubic crystal.;  
 RL Acta Crystallogr. B 47:127-136(1991).  
 RN [10]  
 RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).  
 RA Gao J.-S., Wan Z.-L., Chang W.-R., Jiang D.-C.;  
 RT Structure of monomeric porcine DesR-B2 desptapeptide (R26-R30)  
 insulin at 1.65 A resolution.;  
 RL Acta Crystallogr. D 51:507-512(1997).  
 CC [1]  
 CC FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION.;  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS, AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC [2]  
 CC SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC [3]  
 CC SUBCELLULAR LOCATION: Secreted.  
 CC [4]  
 CC SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC [5]  
 CC DATABASE: NAME-Protein Spotlight;  
 CC NOTE-Issue 9 of April 2001;  
 CC WWW \*http://www.expasy.org/spotlight/articles/spt1r009.html\*.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib.ch](mailto:license@sib.ch)).  
 CC  
 DR EMBL: AF064555; AAC77920.1; ALT\_INIT.  
 DR FMBL: AY044828; AAL69550.1;  
 DR PCB: 3INS; 09-JAN-89.  
 DR PDB: 4INS; 31-JUL-94.  
 DR PDB: 6INS; 31-JAN-94.  
 DR PDB: 7INS; 31-JAN-94.  
 DR PDB: 9INS; 15-OCT-91.  
 DR PDB: 11ZB; 15-OCT-91.  
 DR PDB: 112B; 15-OCT-91.  
 DR PDB: 2TC1; 29-JAN-96.

DR PDB: 1MPJ; 29-JAN-96.  
 DR PDB: 3MTH; 29-JAN-96.  
 DR PDB: 1DEI; 16-JUN-97.  
 DR PDB: 1SDB; 01-APR-98.  
 DR PDB: 1WAV; 28-FEB-97.  
 DR PDB: 1ZEI; 16-FEB-99.  
 DR PDB: 1ZNI; 28-JAN-98.  
 DR PDB: 1ZNJ; 28-JAN-98.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; insulin; 1.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00362; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism; Signal; 3b structure.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54 INSULIN A CHAIN.  
 FT PROPEP 57 85 C PEPTIDE.  
 FT CHAIN 86 108 INSULIN A CHAIN.  
 FT CHAIN 91 94 INTERCHAIN.  
 FT DISULFID 93 98 INTERCHAIN.  
 FT DISULFID 26 45  
 FT HELIX 48 48  
 FT STRAND 89 94  
 FT HELIX 100 106  
 FT STRAND 107 107  
 SQ SEQUENCE 108 AA: CB4491M429R5REBE CRC64;  
 Query Match 89.5%; Score 263; DB 1; Length 108;  
 Best Local Similarity 60.7%; Prod No. 19c-25;  
 Matches 51; Conservative 0; Mismatches 1; Indels 32; Gaps 1;  
 QY 1 FVNQHDSGSHLVEALYLVCGERGFFYTPKT-- ..... 40  
 DB 25 FVNQHGSHLVEALYLVCGERGFFYTPKARAEAFNPQAGAVELGGIGLQALALEGPP 84  
 QY 31 --RGIVEGCTSCSLYQLENYCN 52  
 DB 85 QRKGIVEGCTSCSLYQLENYCN 106  
 RESULT: 12  
 INS\_RAHIT  
 ID INS\_KABIT STANDARD; PRT: 110 AA.  
 AC P01311;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Insulin precursor.  
 GN INS.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus  
 OX NCBI\_TaxID:9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-New Zealand white; TISSUE-Pancreas;  
 RX MEDLINE 94179230; PubMed 8112571;  
 RA Devaskar S.U., Giddings S.J., Rajakumar P.A., Carnaghi L.R.,  
 RA Menon R.K., Zahm D.S.;  
 RT Insulin gene expression and insulin synthesis in mammalian neuronal  
 RT cells.;  
 RL J. Biol. Chem. 269:8445-8454(1994).  
 RN [2]  
 RP SEQUENCE OF 25-54 AND 90-110.  
 RX MEDLINE 66160119; PubMed 5949593;  
 RA Smith L.F.;  
 RT Species variation in the amino acid sequence of insulin.;  
 RL Am. J. Med. 40:662-666(1966).  
 RN [3]  
 RP SEQUENCE OF 56-110 FROM N.A.  
 RA Giddings S.J., Carnaghi L.R., Devaskar S.U.;  
 RL Submitted (APR-1991) to the EMBL/Genbank/DBJ databases  
 CC [4]  
 CC FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT

```

CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETEROIDIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U03610; AAA17033.1;
CC EMBL: M61154; AAA1543.1;
CC PIR: A53448; INRA
CC HSSP: P01308; ILYM.
CC InterPro: IPR004825; Ins/IGF/relax.
CC Pfam: PF00049; Insulin; 1.
CC SMART: SM00078; IIGF; 1.
CC PRISM: PR00277; INSULINB.
CC Kegg: K04490; Hormone; Glucose metabolism; Signal.
CC PROSITE: PS00262; INSULIN; 1.
CC Insulin family; Hormone; Glucose metabolism; Signal.
CC SIGNAL 1 24 INSULIN B CHAIN.
CC CHAIN 25 54 INSULIN B CHAIN.
CC PROPEP 57 87 C PEPTIDE.
CC CHAIN 90 110 INSULIN A CHAIN.
CC DISULFID 31 96 INTERCHAIN (BY SIMILARITY).
CC DISULFID 43 109 INTERCHAIN (BY SIMILARITY).
CC DISULFID 95 100 BY SIMILARITY.
CC CONFLICT 83 83 E -> Y (N REF. 3).
CC SEQUENCE 110 AA; 11838 MW; 8202975885077FA8 CRC64;

Query Match 89.5%; Score 263; DB 1; Length 110;
Best Local Similarity 59.3%; Pred. No. 1,9e-25;
Matches 51; Conservative 1; Mismatches 0; Indels 34; Gaps 1;

QY 1 FVNHLCGSHLVLYVCGRGFFYPTKT--- .. ..... 40
DB 25 FVNHLCGSHLVLYVCGRGFFYPTKSRREVEEQGGVGLGGGAGAGLPOPIALEM 84
QY 31 ---RGIVEQCCTSCSYOLENYCN 52
DB 85 ALOKRGIVEQCCTSCSYOLENYCN 110

RESULT 14
INS_SPETR
ID INS_SPETR STANDARD; PRT: 110 AA.
AC Q91X13.
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Insulin precursor.
GN INS.
OS Spermophilus tridecemlineatus (Thirteen-lined ground squirrel).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
OC Spermophilus.
OX NCBI_TaxID=43179;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Pancreas;
RA Tredrea M.M., Buck M.J., Guhaniyoyi J., Squire T.L., Andrews M.T.;
RT "Regulation of PDK4 expression in a hibernating mammal.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETEROIDIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO

```

```

CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AY038604; AAK72558.1;
CC HSSP: P01308; ILNP.
CC InterPro: IPR004825; Ins/IGF/relax.
CC Pfam: PF00049; Insulin; 1.
CC PRISM: PR00277; INSULINB.
CC SMART: SM00078; IIGF; 1.
CC PROSITE: PS00262; INSULIN; 1.
CC Insulin family; Hormone; Glucose metabolism; Signal.
CC SIGNAL 1 24 BY SIMILARITY.
CC CHAIN 25 54 INSULIN B CHAIN.
CC PROPEP 57 87 C PEPTIDE.
CC CHAIN 90 110 INSULIN A CHAIN.
CC DISULFID 31 96 INTERCHAIN (BY SIMILARITY).
CC DISULFID 43 109 INTERCHAIN (BY SIMILARITY).
CC DISULFID 95 100 BY SIMILARITY.
CC SEQUENCE 110 AA; 12004 MW; 4511768D6622BEE5 CRC64;

Query Match 89.5%; Score 263; DB 1; Length 110;
Best Local Similarity 59.3%; Pred. No. 1,9e-25;
Matches 51; Conservative 1; Mismatches 0; Indels 34; Gaps 1;

QY 1 FVNHLCGSHLVLYVCGRGFFYPTKT--- .. ..... 30
DB 25 FVNHLCGSHLVLYVCGRGFFYPTKSRREVEEQGGVGLGGGAGAGLPOPIALEM 84
QY 31 ---RGIVEQCCTSCSYOLENYCN 52
DB 85 ALOKRGIVEQCCTSCSYOLENYCN 110

RESULT 14
INS_FELCA
ID INS_FELCA STANDARD; PRT: 51 AA.
AC P06306.
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 46, Last annotation update)
DE Insulin.
GN INS.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE.
RA MEDLINE=86214076; PubMed=3518635;
RA Halliden G., Gafvelin G., Mutt V., Joernvall H.;
RT "Characterization of cat insulin.";
RL Arch. Biochem. Biophys. 247:20-27(1986)
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETEROIDIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC PIR: A01588; INCT.
DR HSSP: P01317; IAPH.
DR InterPro: IPR004825; Ins/IGF/relax.
DR PRINTS: PR00277; INSULINB.

```

```

DR SMART: SM00078; IIGF: 1.
DR PROSITE: PS00262; INSULIN: 1.
KW Insulin family; Hormone: Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA: 5745 MW: 9007H5096A0A70DD CRC64:
Query Match 89.1%; Score 262.5; DB 1; Length 51;
Best Local Similarity 90.4%; Pred. No. 2.6e-25;
Matches 47; Conservative 2; Mismatches 1; Gaps 1;
1 FVNQHLCGSHLVEAALYLVGGERGFYTPK1----- 30
111.1111111111111111111111111111111111
25 FVNQHLCGSHLVEAALYLVGGERGFYTPKAPREVEDLVVRDELACAFGEGLQIPIALEG 84
111.1111111111111111111111111111111111
41 --- RELVQVPTISGLSYGLFNYGN 52
1111111111111111111111111111111111
45 ALQKKLVQVPTISGLSYGLFNYGN 110
DL

RESULT 15
INS_CANFA STANDARD: PRT: 110 AA
AC P01321:
DT 21-JUL-1986 (Rel. 01, Created)
DI 21-JUL-1986 (Rel. 01, Last sequence update)
DI 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin precursor.
GN INS.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Mammalia; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID:9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:83109071; PubMed:6296342;
RA Kwok S.C.M., Chan S.J., Steiner D.F.;
RT Cloning and nucleotide sequence analysis of the dog insulin gene.
RT Coded amino acid sequence of canine preproinsulin predicts an
RT additional C-peptide fragment.*;
RL J. Biol. Chem. 258:2357-2363(1983).
[2]
RP SEQUENCE OF 25-54 AND 90-110.
RX MEDLINE:66160114; PubMed:5445953;
RA Smith L.F.;
RT Species variation in the amino acid sequence of insulin.*;
Am. J. Med. 40:662-666(1966).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLUCOGEN SYNTHESIS IN LIVER.
CC -2- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -3- SUBCELLULAR LOCATION: Secreted.
CC -4- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC .....
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC .....
DR EMBL: V00179; CAA23475.1;
DR PIR: A92413; IPDG.
DR HSSP: P01317; IAPH.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF: 1.
DR PROSITE: PS00262; INSULIN: 1.
KW Insulin family; Hormone: Glucose metabolism; Signal.
FT SIGNAL 1 24

```

Search completed: September 15, 2003, 12:01:40  
Job time : 4.84588 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 11:54:30 : Search time 20.3154 seconds  
(without alignments)  
849,520 Million cell updates/sec

Title: US-09-423-100-5

Perfect score: 294

Sequence: 1 FVNHLCGSHLVEALYLVCGR.....IVEQCCTSIQSLYLENYCN 52

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 20625 seqs, 26465604 residues

Total number of hits: 31 (displayed: 30, sorted by score)

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: SP_archaea.*
2: SP_bacteria.*
3: SP_fungi.*
4: SP_human.*
5: SP_invertebrate.*
6: SP_mammal.*
7: SP_mhc.*
8: SP_organelle.*
9: SP_plage.*
10: SP_plant.*
11: SP_rodent.*
12: SP_virus.*
13: SP_vertebrate.*
14: SP_unclassified.*
15: SP_virus.*
16: SP_bacterioid.*
17: SP_archaeap.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	267	96.8	110	6	Q8HXV2 pongo pygma
2	263	89.5	110	11	Q91X13 spermophilu
3	251	85.4	110	6	Q8WNR6 felis silve
4	219.5	74.7	106	14	Q91RQ7 rana pipie
5	201.5	68.5	110	13	Q98TA8 pantodon bu
6	195.5	66.5	108	13	Q9DBF5 brachydanio
7	195.5	66.5	108	13	Q90ZM4 catia catia
8	195	66.3	111	13	Q98160 chitara chi
9	193.5	65.8	110	13	Q90ZY1 hiodon atos
10	181.5	65.1	111	13	Q98TA7 osteoglossu
11	181.5	63.8	87	13	Q98TA9 gnathonemus
12	186.5	63.4	108	13	Q98TB1 catostomus
13	185.5	63.1	91	13	Q98TB2 amphiolites
14	146	49.7	65	6	Q8H281 gorilla gor
15	146	49.7	65	6	Q8H280 pongo pygma
16	144	49.0	207	13	Q90XD0 cyprinus ca

17	144	49.0	215	13	073721	tilapia sp.
18	143.5	48.8	132	13	Q8AV14	petromyzon
19	141.5	48.1	159	13	Q93607	paralichthy
20	141.5	48.1	182	13	073720	oreochromis
21	141.5	48.1	182	13	042289	oreochromis
22	141.5	48.1	182	13	P79824	oreochromis
23	141.5	48.1	185	13	057436	paralichthy
24	141.5	48.1	186	13	093527	paralichthy
25	141	48.0	185	13	Q9V157	acanthopagr
26	140.5	47.8	116	13	Q9V161	oncorhynch
27	140.5	47.8	117	13	Q9V176	salmo salar
28	140.5	47.8	145	13	Q95475	salmo salar
29	140.5	47.8	145	13	Q91231	oncorhynch
30	140.5	47.8	155	13	Q91162	oncorhynch
31	140.5	47.8	161	13	Q91230	oncorhynch
32	140.5	47.8	184	13	042336	myoxocephal
33	140.5	47.8	188	13	P81268	oncorhynch
34	140.5	47.8	188	13	Q91965	oncorhynch
35	140	47.6	210	14	Q91443	squalus aca
36	139	47.3	212	14	Q81164	brachydanio
37	139	47.3	216	13	042429	latas calca
38	138.5	47.1	124	6	Q8M115	sus scrofa
39	138.5	47.1	134	13	Q93380	melagadis g
40	137.5	46.8	62	13	Q91AA0	carassius a
41	137.5	46.8	117	13	Q91914	ctenopharyn
42	137.5	46.8	149	6	Q9YX4	bos indicus
43	137.5	46.8	163	13	Q90VV9	brachydanio
44	137.5	46.8	161	13	Q9PK42	carassius a
45	137.5	46.8	161	13	Q98SR6	megalobrama

#### ALIGNMENTS

##### RESULT 1

```

Q8HXV2
ID   Q8HXV2          PRELIMINARY:          PRT:   110 AA.
AC   Q8HXV2:
DT   01-MAR-2003 (TREMBLrel. 23, Created)
DT   01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE   01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE   Insulin precursor.
GN   INS.
OS   Pongo pygmaeus (Gorakanan).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Hominidae; Pongo.
OX   NCBI_TaxID:9600;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stead J.D.H., Jeffreys A.J.,
RT   "Haplotype diversity at the insulin region."
RL   Submitted (Jul-2002) to the EMBL/GenBank/DBJ databases.
DR   EMBL: AY137503; AAN06917;
SQ   SEQUENCE 110 AA: 12038 MW: 2202832B94F520F8 CRC64;

```

Query Match 90.8% Score 267; DB 6; Length 110;  
Best Local Similarity 60.5%; Pred. No. 16-28;  
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 1 FVNHLCGSHLVEALYLVCGRGFYTPKPT----- 30

DB 25 FVNHLCGSHLVEALYLVCGRGFYTPKPTKRAFLQVQVLELGGIPAGSLQPLALEG 84

QY 31 -----RGIVEQCCTSIQSLYLENYCN 52

DB 85 SLQKRGIVEQCCTSIQSLYLENYCN 110

##### RESULT 2

```

Q91X13
ID   Q91X13          PRELIMINARY:          PRT:   110 AA.
AC   Q91X13:
DT   01-DEC-2001 (TREMBLrel. 19, Created)

```



DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
 DE Insulin.  
 OS *Spermophilus tridecemlineatus* (Thirteen-lined ground squirrel).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;  
 CC Spermophilus.  
 OX NCBI\_TaxID=43179;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=pancreas;  
 RA Tiedora M.M., Buck M.J., Ghanayogi S., Squire T.L., Andrews M.T.;  
 RT "Regulation of PK4 expression in a hibernating mammal";  
 RL Submitted (JUN-2001) to the EMBL/Genbank/DDBJ databases.  
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL: AY048694; AAK72568.1;  
 DR HSP: P01308; 1; NDB;  
 DR InterPro: IPR04825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1;  
 DR PRINTS: PK00277; INSULIN;  
 DR SMART: SM00078; IIGF; 1;  
 DR PROSITE: PS00262; INSULIN; 1;  
 SQ SEQUENCE 110 AA: 12004 MW: 451176866622BEE5 CRC64:

Query Match 89.5%; Score 263; DB 11; Length 110;  
 Best Local Similarity 59.3%; Pred. No. 4,6e-26;  
 Matches 59; Conservative 1; Mismatches 0; Indels 14; Gaps 1;  
 QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKT----- 30  
 DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKSRREVFQGGVGLGPGAGLQPLALEM 84  
 QY 31 ----RGIVEQCCTSLCSLYOLENYCN 52  
 DB 85 ALQKRGIVEQCCTSLCSLYOLENYCN 110

RESULT 3  
 Q8WNW6  
 ID Q8WNW6 PRELIMINARY: PRT: 110 AA.  
 AC Q8WNW6;  
 DT 01-MAR-2002 (TRENBLrel. 20, Created)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 21, Last annotation update)  
 DE Preproinsulin.  
 OS *Felis silvestris catus* (Cat).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
 OX NCBI\_TaxID=9685;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=pancreas;  
 RA Okamoto S., Morimatsu M.;  
 RT "cat insulin";  
 RL Submitted (MAY-2000) to the EMBL/Genbank/DDBJ databases.  
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL: AB043535; BAB94110.1;  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1;  
 DR PRINTS: PK00277; INSULIN;  
 DR SMART: SM00078; IIGF; 1;  
 DR PROSITE: PS00262; INSULIN; 1;  
 SQ SEQUENCE 110 AA: 12069 MW: 95FB6E170C7BEC44 CRC64:

Query Match 85.4%; Score 251; DB 6; Length 110;  
 Best Local Similarity 55.8%; Pred. No. 1.6e-26;  
 Matches 48; Conservative 2; Mismatches 2; Indels 34; Gaps 1;  
 QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKT----- 30  
 DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKARREAEDLOCKDAELGEPAGGGLQPSALEA 84

QY 41 ----RGIVEQCCTSLCSLYOLENYCN 52  
 DB 85 PLQKRGIVEQCCTSLCSLYOLENYCN 110  
 RESULT 4  
 Q91BQ7  
 ID Q91BQ7 PRELIMINARY: PRT: 106 AA.  
 AC Q91BQ7;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Preproinsulin.  
 OS *Rana pipiens* (Northern leopard frog).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Amphibia; Batrachia; Anura; Nothofactia; Ranidae; Rana.  
 OX NCBI\_TaxID=8404;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=20362507; PubMed 10618274;  
 RA Irwin D.M., Sivarajah P.;  
 RT "Proinsulin cDNAs from the leopard frog, *Rana pipiens*: evolution of  
 RT proinsulin processing";  
 RL Comp. Biochem. Physiol. 125B:405-410(2000).  
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL: AF227187; AAF67285.1;  
 DR HSP: P01315; 1; NDB;  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1;  
 DR PRINTS: PK00277; INSULIN;  
 DR SMART: SM00078; IIGF; 1;  
 DR PROSITE: PS00262; INSULIN; 1;  
 SQ SEQUENCE 106 AA: 12183 MW: 3A870EFC70217F92 CRC64:

Query Match 74.7%; Score 219; DB 5; DB 13; Length 106;  
 Best Local Similarity 49.4%; Pred. No. 1e-22;  
 Matches 41; Conservative 7; Mismatches 3; Indels 31; Gaps 1;  
 QY 1 FVNHLCGSHLVEALYLVCGERGFFYTPKT----- 31  
 DB 24 FQNGVLCGSHLVEALYMWCGRGFFYSPRSKRDLQPLVNSLGGSEIDEMQVSOAFQKR 83  
 QY 42 -GIVEQCCTSLCSLYOLENYCN 52  
 DB 84 KPGIVEQCCHNTTSLYOLENYCN 106  
 RESULT 5  
 Q9RTAR  
 ID Q9RTAR PRELIMINARY: PRT: 110 AA.  
 AC Q9RTAR;  
 DT 01-JUN-2001 (TRENBLrel. 17, Created)  
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Preproinsulin.  
 OS *Pantodon buchholzi* (Butterflyfish).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
 CC Osteoglossiformes; Pantodontidae; Pantodon.  
 OX NCBI\_TaxID=8276;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=21203577; PubMed 11304171;  
 RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;  
 RT "Molecular cloning of preproinsulin cDNAs from several  
 RT osteoglossomorphs and a cyprinid";  
 RL Mol. Cell. Endocrinol. 174:51-58(2001).  
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL: AF199588; AAK28712.1;  
 DR HSP: P01308; 1; NDB.

```

DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
SQ SEQUENCE 110 AA; 12324 MW; BDCCD659 472E06 CRC64;

Query Match
Best Local Similarity 68.5%; Score 201.5; DB 13; Length 110;
Matches 37; Conservative 8; Mismatches 5; Indels 35; Gaps 1;

QY 3 NMLGCSHLVVALYVCGERGFYTPK-----T 30
Db 26 SORLGCSHLVDALYVCGERGFYTPKRDVPLLGFLPKSAQTEVAFAFKDHAEC;RK 85
QY 31 ---RGIVEQCCTSGISLYOLENYCN 52
Db 86 KVRKGVVCGCHHPNIFELQNYCN 108

RESULT 4
Q902N4 PRELIMINARY; PRT: 108 AA.
AC Q902N4
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TRENBLrel. 23, Last annotation update)
DE Insulin precursor.
GN INS.
OS Brachydanio rerio (zebrafish) (Danio rerio);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxId=7955;
RN [1]
RX MEDLINE:94425190; PubMed:10495291;
RA Argenton F., Zecchin E., Bortolussi M.;
RT "Early appearance of pancreatic hormone-expressing cells in the
RT zebrafish embryo."
RC Mech. Dev. 87:217-221(1999).
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL: AJ237750; CAC20109.1;
DR HSSP: 901308; 11PH.
DR ZFIN: ZDB-GENE-980526-110; ins.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00277; INSULIN.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
KW Signal.
FT SIGNAL. 1 24 POTENTIAL.
FT CHAIN. 24 53 INSULIN A CHAIN.
FT CHAIN. 86 108 INSULIN B CHAIN.
SQ SEQUENCE 108 AA; 11904 MW; 3105289E7A06D25 CRC64;

Query Match
Best Local Similarity 66.5%; Score 195.5; DB 13; Length 108;
Matches 37; Conservative 5; Mismatches 7; Indels 43; Gaps 1;

QY 4 QHLCGSHLVVALYVCGERGFYTPK-----T 30
Db 27 QHLCGSHLVDALYVCGTFGFYTPKRDVPLLGFLPKSAQTEVAFAFKDHAEC;RK 86
QY 31 RGIVEQCCTSGISLYOLENYCN 52
Db 87 RGIVEQCCHKPCSIFELQNYCN 108

RESULT 7
Q902N4 PRELIMINARY; PRT: 108 AA.
AC Q902N4
DT 01-MAR-2001 (TRENBLrel. 19, Created)
DT 01-MAR-2001 (TRENBLrel. 19, Last sequence update)
DE 01-MAR-2001 (TRENBLrel. 23, Last annotation update)
DE Preproinsulin.
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Catla.
OX NCBI_TaxId=72446;
RN [1]
RX MEDLINE:94425190; PubMed:10495291;
RA Argenton F., Zecchin E., Bortolussi M.;
RT "Early appearance of pancreatic hormone-expressing cells in the
RT zebrafish embryo."
RC Mech. Dev. 87:217-221(1999).
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL: AF199486; AAK28710.1;
DR HSSP: P01308; 11PH.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
FT NON_TER. 111 111
SQ SEQUENCE 111 AA; 12483 MW; 247CA4431376329F CRC64;

Query Match
Best Local Similarity 66.3%; Score 195; DB 13; Length 111;
Matches 38; Conservative 3; Mismatches 9; Indels 36; Gaps 1;

```

```

DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DE 01-MAR-2001 (TRENBLrel. 23, Last annotation update)
DE Preproinsulin.
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Catla.
OX NCBI_TaxId=72446;
RN [1]
RX MEDLINE:94425190; PubMed:10495291;
RA Argenton F., Zecchin E., Bortolussi M.;
RT "Early appearance of pancreatic hormone-expressing cells in the
RT zebrafish embryo."
RC Mech. Dev. 87:217-221(1999).
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL: AF199486; AAK28710.1;
DR HSSP: P01308; 11PH.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
FT NON_TER. 111 111
SQ SEQUENCE 111 AA; 12483 MW; 247CA4431376329F CRC64;

Query Match
Best Local Similarity 66.5%; Score 195.5; DB 13; Length 108;
Matches 37; Conservative 5; Mismatches 7; Indels 33; Gaps 1;

QY 4 QHLCGSHLVVALYVCGERGFYTPK-----T 30
Db 27 QHLCGSHLVDALYVCGTFGFYTPKRDVPLLGFLPKSAQTEVAFAFKDHAEC;RK 86
QY 31 RGIVEQCCTSGISLYOLENYCN 52
Db 87 RGIVEQCCHKPCSIFELQNYCN 108

RESULT 8
Q98TB0 PRELIMINARY; PRT: 111 AA.
AC Q98TB0
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DE 01-MAR-2001 (TRENBLrel. 23, Last annotation update)
DE Preproinsulin (Fragment).
OS Chitala chitala (crown knifefish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
OC Osteoglossiformes; Notopteridae; Chitala.
OX NCBI_TaxId=112153;
RN [1]
RX MEDLINE:21203577; PubMed:11361171;
RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
RT osteoglossomorphs and a cyprinid."
RC Mol. Cell. Endocrinol. 174:51-58(2001).
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL: AF199486; AAK28710.1;
DR HSSP: P01308; 11PH.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
FT NON_TER. 111 111
SQ SEQUENCE 111 AA; 12483 MW; 247CA4431376329F CRC64;

Query Match
Best Local Similarity 66.3%; Score 195; DB 13; Length 111;
Matches 38; Conservative 3; Mismatches 9; Indels 36; Gaps 1;

```

```

QY 3 NOHLGCSHLVLEALYVCGERGFFYTPK----- 29
DB 26 NOHLGCSHLVLEALYVCGERGFFYTPKMKRUAEPILGLSPKSGLENEVDYPPKDGDL 85
QY 30 ---TRGIVEQCCSTICSLYLENYCN 52
DB 86 VVKRGIVEQCCHRCNIFDNOYCN 111

RESULT 9
ID Q98TA7 PRELIMINARY: PRT: 111 AA.
AC Q98TA7
DT 01-JUN-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Preproinsulin (Fragment)
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha
OC Osteoglossiformes; Blenniidae; Blenni
OX NCBI_TaxID:54904
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:21203577; PubMed:11306171;
RA Al Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
osteoglossomorphs and a cyprinid."
RL Mol. Cell. Endocrinol. 174:51-58(2001).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
DR EMBL: AF282408; AAK5684.1;
DR HSSP: P01308; IHP.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
FT NON_TER 110 111
SQ SEQUENCE 110 AA: 12343 MW: 80800770452606 CRC64:
Query Match 65.8%; Score 191.5; DB 13; Length 110;
Best Local Similarity 42.4%; Pred. No. 2, 1e-18;
Matches 36; Conservative 7; Mismatches 7; Indels 45; Gaps 1;

QY 3 NOHLGCSHLVLEALYVCGERGFFYTPK----- 30
DB 26 NOHLGCSHLVLEALYVCGERGFFYTPKMKRUAEPILGLSPKSGLENEVDYPPKDGDL 85
QY 31 ---RGIVEQCCSTICSLYLENYCN 52
DB 86 VVKRGIVEQCCHRCNIFDNOYCN 111

RESULT 13
Q98TA7 PRELIMINARY: PRT: 111 AA.
AC Q98TA7
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 20, Last annotation update)
DE Preproinsulin (Fragment)
OS Osteoglossum bicirrhosum (silver arawaka).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha
OC Osteoglossiformes; Osteoglossidae; Osteoglossum.
OX NCBI_TaxID:109271;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:21203577; PubMed:11306171;
RA Al Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
osteoglossomorphs and a cyprinid."

```

```

RL Mol. Cell. Endocrinol. 174:51-58(2001).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL: AF199589; AAK2871.1;
DR HSSP: P01315; IHPJ.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00277; INSULINB.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
FT NON_TER 111 111
SQ SEQUENCE 111 AA: 12491 MW: 8098190204866823 CRC64:
Query Match 65.1%; Score 191.5; DB 13; Length 111;
Best Local Similarity 41.2%; Pred. No. 2, 1e-18;
Matches 35; Conservative 10; Mismatches 5; Indels 35; Gaps 1;

QY 4 OHLCGSHLVEALYVCGERGFFYTPK----- 30
DB 27 SOHLGCSHLVLEALYVCGERGFFYTPKMKRUAEPILGLSPKSGLENEVDYPPKDGDL 86
QY 31 ---RGIVEQCCSTICSLYLENYCN 52
DB 87 VVKRGIVEQCCHRCNIFDNOYCN 111

RESULT 11
Q98TA7 PRELIMINARY: PRT: 67 AA.
AC Q98TA7
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Preproinsulin (Fragment)
OS Gnathionemus petrusii.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha
OC Osteoglossiformes; Mormyridae; Gnathionemus.
OX NCBI_TaxID:42645;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:21203577; PubMed:11306171;
RA Al Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
osteoglossomorphs and a cyprinid."
RL Mol. Cell. Endocrinol. 174:51-58(2001).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
DR EMBL: AF199587; AAK2871.1;
DR HSSP: P01308; IHPJ.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
FT NON_TER 87 87
SQ SEQUENCE 87 AA: 8874 MW: 884880350245385 CRC64:
Query Match 63.8%; Score 187.5; DB 13; Length 87;
Best Local Similarity 42.9%; Pred. No. 5, 8e-18;
Matches 36; Conservative 5; Mismatches 8; Indels 35; Gaps 1;

QY 4 OHLCGSHLVEALYVCGERGFFYTPK----- 30
DB 4 OHLCGSHLVEALYVCGERGFFYTPKMKRUAEPILGLSPKSGLENEVDYPPKDGDL 84
QY 31 ---RGIVEQCCSTICSLYLENYCN 52
DB 64 VVKRGIVEQCCHRCNIFDNOYCN 87

RESULT 12
Q98TB1

```

```

ID Q98TB1 PRELIMINARY: PRT: 108 A.
AC Q98TB1:
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Preproinsulin (Fragment).
OS Catostomus commersoni (White sucker).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Catostomidae; Catostomus.
OX NCBI_TaxID=7971;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 2103577; PubMed: 11306171;
RA Al-Mabrouk A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
RT osteoglossomorphs and a cyprinid."
RL Mol. Cell. Endocrinol. 174:51-56 (2000).
CC -1. SIMILARITY: BELONGS TO THE INSULIN/F/RELAXIN FAMILY.
DE EMBL: AF194585; AAK28709.1;
DR HSSP: P01308; 11PF.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
FT NON_TER 108
FT SEQUENCE 108 AA: 1187; MW: 8426310696F9AFC8 CRC64;

Query Match: 63.4%; Score 186.5; DB 13; Length 108;
Best Local Similarity 41.9%; Pred. No. 9, 80-18;
Matches 36; Conservative 4; Mismatches 9; Indels 33; Gaps 1;

QY 4 QHLCGSHLVVYALVCGGRCFFYTPK .....T 30
DB 27 QHLCGSHLVVYALVCGGRCFFYTPK .....T 30
QY 31 KQIVEQCCTSCSLYULENYCN 52
DB 87 KQIVEQCCTSCSLYULENYCN 108

RESULT 13
Q98TB2 PRELIMINARY: PRT: 91 AA.
AC Q98TB2:
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Preproinsulin (Fragment).
OS Ambloplites rupestris (Rock bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoleptoostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Percoidae;
OC Centrarchidae; Ambloplites.
OX NCBI_TaxID=109273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 2103577; PubMed: 11306171;
RA Al-Mabrouk A.A., Irwin D.M., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNA from the rock bass."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -1. SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1. SIMILARITY: BELONGS TO THE INSULIN/F/RELAXIN FAMILY.
DE EMBL: AF194584; AAK28708.1;
DR HSSP: P01308; 11PF.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
FT NON_TER 91
FT SEQUENCE 91 AA: 10100 MW: 886C8B256DC69D39 CRC64;

```

```

Query Match: 63.18; Score 185.5; DB 13; Length 91;
Best Local Similarity 40.9%; Pred. No. 1, 11e-17;
Matches 36; Conservative 5; Mismatches 8; Indels 39; Gaps 1;

QY 4 QHLCGSHLVVYALVCGGRCFFYTPK ..... 29
DB 4 QHLCGSHLVVYALVCGGRCFFYTPK ..... 29
QY 30 ---- TKGIVEQCCTSCSLYULENYCN 52
DB 64 MEMVKRGSLVEQCCHHPCNIFDLRYCN 91

RESULT 14
Q98TB1 PRELIMINARY: PRT: 65 AA.
AC Q98TB1:
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Insulin (Fragment).
OS Gorilla gorilla (gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Gorilla.
OX NCBI_TaxID=9593;
RN [1]
RP SEQUENCE FROM N.A.
RX "Molecular evolution in higher primates: gene specific and organism
RT specific characteristics."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DE EMBL: AY092023; AAM76640.1;
FT NON_TER 1
FT NON_TER 65
FT SEQUENCE 65 AA: 5920 MW: B772017FD8HCAHEA CRC64;

Query Match: 49.7%; Score 146; DB 6; Length 65;
Best Local Similarity 47.7%; Pred. No. 2e-12;
Matches 31; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 7 CGSHLVVYALVCGGRCFFYTPK .....RG 32
DB 3 CGSHLVVYALVCGGRCFFYTPK ..... 32
QY 33 IVEQC 37
DB 61 IVEQC 65

RESULT 15
Q98TB0 PRELIMINARY: PRT: 65 AA.
AC Q98TB0:
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Insulin (Fragment).
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP SEQUENCE FROM N.A.
RX "Molecular evolution in higher primates: gene specific and organism
RT specific characteristics."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DE EMBL: AY092024; AAM76641.1;
FT NON_TER 1
FT NON_TER 65
FT SEQUENCE 65 AA: 6920 MW: B772017FD8HCAHEA CRC64;

Query Match: 49.7%; Score 146; DB 6; Length 65;

```

```

Best Local Similarity 47.7%; Pred. No. 2e-12;
Matches 31; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 7 CGSHLVEALYLCGERGFFYTPKT-----RG 32
   |||||||'|||||'|||||
Db 1 CGSHLVEALYLCGERGFFYTPKTRREADLQVQVVELGGGPGAGSIQPLALEGSLKRG 60
   |||
QY 33 IVFQC 37
   ||||
Db 61 IVFQC 65

```

Search completed: September 15, 2003, 12:03:30  
Job time : 21.3154 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 12:31:20 ; Search time 31.642 Seconds  
(without alignments)  
536.746 Million cell updates/sec

Title: US-09-423-100-6  
Perfect score: 587  
Sequence: 1 MPTIPLSRFLDNALRAHR.....IVEOCTSIQSLYLENYCN 107

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 res.lues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	587	100.0	107	20 AAY42860	hGH-mini-proinsulin
2	555.5	94.6	150	20 AAY42861	Chimeric protein,
3	315.5	53.7	116	17 AAR98897	SOD-proinsulin hyb
4	304	51.8	63	15 AAR68900	Human pro-insulin
5	304	51.8	117	17 AAR98896	SOD-proinsulin hyb
6	302.5	51.5	137	16 AAR71692	Mating factor alph
7	299	50.9	56	15 AAR68901	Human pro-insulin
8	299	50.9	56	16 AAR78665	Proinsulin sequenc
9	299	50.9	96	15 AAR68899	Human pro-insulin

10	299	50.9	96	16 AAR78662	Fusion protein con
11	299	50.9	145	16 AAR71694	Mating factor alph
12	299	50.9	146	16 AAR71695	Mating factor alph
13	294	50.1	52	20 AAY42859	Human insulin prec
14	293	49.9	57	11 AAR04582	Proinsulin analogu
15	288.5	49.1	160	16 AAR79056	Glycosylphosphatid
16	287	48.9	52	12 AAR11899	Example of human i
17	287	48.9	65	19 AAW47365	Preproinsulin 1.
18	287	48.9	138	1 AAR87086	pK142 modified in
19	284.5	48.5	58	1 AAR96047	Modified pJ859 mod
20	284.5	48.5	59	1 AAR96048	Modified pJ859 mod
21	284.5	48.5	65	1 AAR88188	N-terminally exten
22	284.5	48.5	109	10 AAP94645	Amino acids encode
23	284.5	48.5	123	18 AAW19240	EEAEPK-M13 insulin
24	284.5	48.5	123	19 AAW69160	DNA construct pAK7
25	284.5	48.5	124	19 AAW78751	PAK855 protein seq
26	284.5	48.5	124	24 ABB82578	Synthetic leader p
27	284.5	48.5	124	24 ABB82578	Insulin precursor
28	284.5	48.5	125	18 AAP55059	EEAEPK-M15 insulin
29	284	48.4	138	10 AAP94643	Amino acids encode
30	284	48.4	138	17 AAW04890	S. cerevisiae MF a
31	284	48.4	140	16 AAR71693	Mating factor alph
32	284	48.4	140	16 AAR71690	Mating factor alph
33	283.5	48.3	53	1 AAR65883	Di-Arg-(B31-32)-Hu
34	283.5	48.3	53	1 AAW18007	Insul double-chain
35	283.5	48.3	117	11 AAW78752	Protein sequence o
36	283.5	48.3	408	22 AAB30705	A Bacillus pectate
37	282	48.0	94	10 AAP94644	Amino acids encode
38	282	48.0	120	18 AAW19241	EEAEPK-M11 insulin
39	281.5	48.0	60	3 AAP20002	Human proinsulin a
40	281.5	48.0	105	1 AAP94648	Amino acids encode
41	281.5	48.0	153	21 AAY53589	Human proinsulin
42	281	47.9	102	10 AAP94649	Signal-leader-p-in
43	281	47.9	104	16 AAR71684	Yeast signal/leade
44	280.5	47.8	55	8 AAP71019	Sequence encoded b
45	280.5	47.8	89	17 AAR88179	Signal peptide/lea

ALIGNMENTS

RESULT 1  
ID AAY42860 standard; protein; 107 AA.

XX AC AAY42860:  
XX 19-JAN-2000 (first entry)  
XX DE hGH-mini-proinsulin chimeric protein.

XX Insulin; precursor; growth hormone; chaperone; intramolecular;  
KW folding; conformation; chimeric protein; cleavable; recombinant;  
KW production; yield.

XX OS Synthetic.  
XX OS Homo sapiens.

PD 07-OCT-1999.

XX PF 31-MAR-1998; 98WO-CN00052.

XX PR 31-MAR-1998; 98WO-CN00052.

XX (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.

XX WPI; 1999-610839/52.

PT New chimeric proteins containing human growth hormone fragment, used

PT particularly for the production of human insulin -  
 PS Claim 13; Page 30; 46pp; English.  
 XX  
 CC This sequence represents a chimeric protein, hGH-mini-proinsulin.  
 CC This chimeric protein contains an N-terminal fragment of human growth  
 CC hormone (hGH) of the sequence given in AAY42855, a cleavable peptide  
 CC linker (AAY42857), and a human insulin precursor comprising insulin  
 CC A and B chains (AAY42859). The hGH portion of the chimeric protein acts  
 CC as an intramolecular chaperone (IMC) for the insulin precursor,  
 CC enabling it to fold correctly. The cleavable peptide linker has a  
 CC C-terminal Arg residue which enables the hGH portion of the  
 CC chimeric protein to be removed after folding has taken place. Production  
 CC of recombinant human insulin via an hGH-proinsulin chimeric protein can  
 CC provide human insulin with correctly linked cysteine bridges with  
 CC fewer necessary procedural steps, and hence resulting in a higher yield  
 CC of human insulin. The IMC sequences not only protect insulin sequences  
 CC from intracellular degradation by a microorganism host, but also promote  
 CC the folding of the fused insulin precursor, facilitate the solubility of  
 CC the fusion protein and decrease the intermolecular interactions among  
 CC the fusion proteins, thus allowing folding of the fused insulin precursor  
 CC at commercially useful high concentrations. The procedural steps of  
 CC cyanogen bromide cleavage, oxidative sulphitolysis and related  
 CC purification steps can thus be eliminated, along with the use of high  
 CC concentrations of mercaptan or the use of hydrophobic absorbent resins.  
 XX  
 SQ Sequence 107 AA;

Query Match 100.0%; Score 587; DB 20; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 4e-42;  
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEPTIPLSRLEFDNAMLRAHRLHQLAFDITYQEFEEAYIPKEOKYSFLQNPGLTGRPFVNHQ 60  
 DB 1 MEPTIPLSRLEFDNAMLRAHRLHQLAFDITYQEFEEAYIPKEOKYSFLQNPGLTGRPFVNHQ 60

QY 61 LCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYQLENYCN 107  
 DB 61 LCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYQLENYCN 107

RESULT 2  
 AAY42861  
 ID AAY42861 standard; protein: 150 AA.  
 XX  
 AC AAY42861;  
 XX  
 DT 19-JAN-2000 (first entry)  
 DE Chimeric protein, SEQ ID 7.  
 XX  
 KW Insulin; precursor: growth hormone; chaperone: intramolecular;  
 KW folding; conformation: chimeric protein; cleavable; recombinant;  
 KW production; yield.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 PN WO950302-A1.  
 XX  
 PD 07-OCT-1999.  
 XX  
 PF 31-MAR-1998; 98WO-CN00052.  
 XX  
 PR 31-MAR-1998; 98WO-CN00052.  
 XX  
 PA (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.  
 XX  
 PI Gan Z;  
 XX  
 DR WPI: 1999-610839/52.  
 XX  
 PT New chimeric proteins containing human growth hormone fragment, used

PT particularly for the production of human insulin -  
 PS Claim 14; Page 30-31; 46pp; English.  
 XX  
 CC This sequence represents a chimeric protein, which contains an  
 CC N-terminal fragment of human growth hormone (hGH) of the sequence given  
 CC in AAY42856, a cleavable peptide linker (AAY42857), and a human insulin  
 CC precursor comprising insulin A and B chains (AAY42859). The hGH portion  
 CC of the chimeric protein acts as an intramolecular chaperone (IMC) for  
 CC the insulin precursor, enabling it to fold correctly. The cleavable  
 CC peptide linker has a C-terminal Arg residue which enables the hGH portion  
 CC of the chimeric protein to be removed after folding has taken place.  
 CC Production of recombinant human insulin via an hGH-proinsulin chimeric  
 CC protein can provide human insulin with correctly linked cysteine bridges  
 CC with fewer necessary procedural steps, and hence resulting in a higher  
 CC yield of human insulin. The IMC sequences not only protect insulin  
 CC sequences from intracellular degradation by a microorganism host, but  
 CC also promote the folding of the fused insulin precursor, facilitate the  
 CC solubility of the fusion protein and decrease the intermolecular  
 CC interactions among the fusion proteins, thus allowing folding of the  
 CC fused insulin precursor at commercially useful high concentrations. The  
 CC procedural steps of cyanogen bromide cleavage, oxidative sulphitolysis  
 CC and related purification steps can thus be eliminated, along with the use  
 CC of high concentrations of mercaptan or the use of hydrophobic absorbent  
 CC resins.  
 XX  
 SQ Sequence 150 AA;

Query Match 94.6%; Score 555.5; DB 20; Length 150;  
 Best Local Similarity 71.3%; Pred. No. 2.4e-39;  
 Matches 107; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MEPTIPLSRLEFDNAMLRAHRLHQLAFDITYQEFEEAYIPKEOKYSFLQNP----- 49  
 DB 1 MEPTIPLSRLEFDNAMLRAHRLHQLAFDITYQEFEEAYIPKEOKYSFLQNPQTSLSFSSEIP 60

QY 50 -----LGTGPRFVNHQSLHVEALYLVCGER 77  
 DB 61 TPSNREETQOKSNLELLRLISLLLIQSWLEPVLQGTGPRFVNHQSLHVEALYLVCGER 120

QY 78 GFFYTPKTRGIVEQCCTSCISLYQLENYCN 107  
 DB 121 GFFYTPKTRGIVEQCCTSCISLYQLENYCN 150

RESULT 3  
 AAR98897  
 ID AAR98897 standard; Protein: 116 AA.  
 XX  
 AC AAR98897;  
 XX  
 DT 03-FEB-1997 (first entry)  
 DE SOD-proinsulin hybrid polypeptide.  
 XX  
 KW Insulin; proinsulin; hybrid polypeptide; protein folding;  
 KW enzymatic cleavage; cyanogen bromide; sulphitolysis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9620724-A1.  
 XX  
 PD 11-JUL-1996.  
 XX  
 PF 29-DEC-1994; 94WO-US13268.  
 XX  
 PR 29-DEC-1994; 94WO-US13268.  
 PR 10-JAN-1995; 95ZA-0000142.  
 XX  
 PA (BIOT-) BIO-TECHNOLOGY GENERAL CORP.  
 XX  
 PI Gorecki M, Hartman JR, Mendelovitz S;  
 XX







```

CC bridges. Reaction with trypsin and opt. carboxypeptidase B yields
CC correctly folded insulin. The insulin is isolated by absorption on
CC a hydrophobic resin.
XX
XX Sequence 56 AA;
SQ
Query Match 50.9%; Score 299; DB 16; Length 56;
Best Local Similarity 100.0%; Pred. No. 3.2e-18;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 55 RFVNHLCGSHLVEALYLIVGGERGFFYTPKTRGIVEOCCTISCSLYOILENYCN 1-7
Db 4 RFVNHLCGSHLVEALYLIVGGERGFFYTPKTRGIVEOCCTISCSLYOILENYCN 5-
RESULT: 9
AAR68899
ID AAR68899 standard; peptide: 96 AA.
XX
XX AAR68899;
AC
XX
DT 25-MAR-2003 (updated)
DI 02-MAR-1995 (first entry)
XX
XX Human pro-insulin 2.
DE
DE Pro-insulin; A chain; B chain; C-chain; disulphide;
KW mercaptan; chaotropic agent.
XX
XX Homo sapiens.
OS
XX EP600372-A1.
PN
XX 08-JUN-1994.
XX
XX 25-NOV-1993; 93EP-0118993.
XX
XX 02-DEC-1992; 92DE-4240420.
XX (FARH ) HOECHST AG.
XX
XX Gerl M, Ludwig J, Obermeier R, Sabel W;
PI WPI; 1994-177718/22.
XX
XX Prodn. of pro-insulin with direct disulphide bridges by
PT treating recombinant precursor protein with mercaptan in alkali;
PT and in presence of chaotropic agent, then isolation on
PT hydrophobic resin
XX
PS Disclosure; Page 11; 15pp; German.
XX
XX Pro insulin is produced by treating recombinant precursor protein
CC with a mercaptan to produce 2-10 SH residues per Cys residue, in
CC presence of a chaotropic agent and in an medium of pH 10-11,
CC treating the prod. with 3-50 g hydrophobic adsorber resin per 1 g
CC medium of pH 4-7, isolating the adsorbed resin and pro-insulin and
CC desorbing the pro-insulin. This method produces pro-insulin with
CC correctly bonded Cys bridges. Compared with known methods it
CC involves fewer stages (esp. no sulfitolysis or cyanogen bromide
CC cleavage) and overall losses during purification are reduced, i.e.
CC the process is quicker and gives better yields.
CC Sequences of insulin chain A, B and C are given in AAR68895-97.
CC Sequences of pro-insulin 1-4 are given in AAR68898-901.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 96 AA;
SQ
Query Match 50.9%; Score 299; DB 15; Length 96;
Best Local Similarity 100.0%; Pred. No. 5.1e-18;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 55 RFVNHLCGSHLVEALYLIVGGERGFFYTPKTRGIVEOCCTISCSLYOILENYCN 107

```

|||||  
 44 RFVNHLCGSHLVEALVLCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 96

Db

RESULT 10  
 AAR78662

ID AAR78662 standard; protein: 96 AA.

XX AAR78662;

AC 03-APR-1996 (first entry)

XX Fusion protein contg. proinsulin sequence 3.

XX Proinsulin: post-translational modification; recombinant production;  
 protein folding; conformation.

XX Synthetic.

XX Key Location/Qualifiers

FT Region 41..44

FT /label= R2

FT /note= "a peptide of 4 amino acids"

FT Peptide 45..74

FT /label= R1-(B2-B29)-Y

FT /note= "human insulin B-chain"

FT Region 75

FT /label= X

FT Peptide 76..96

FT /label= Gly-(A2-A20)-R3

FT /note= "human insulin A-chain"

XX EP668292-A2.

XX 23-AUG-1995.

XX 09-FEB-1995; 95EP-0101748.

XX 18-FEB-1994; 94DE-4405179.

XX (FAKH) HOECHST AG.

XX Gerl M, Ludwig J, Obermeier R, Sabel W;

XX WPI: 1995-284754/38.

XX Isolation of insulin that is correctly post-translationally  
 processed - by reacting pro-insulin with a mercaptan in the presence  
 of a chaotropic agent and purification after absorption to hydrophobic  
 resin

XX Example 2; Page 8; 16pp; German.

XX The present sequence is that of a fusion protein, produced in E.coli  
 CC which contains an example of a proinsulin molecule corresp.

CC to the general formula R2-R1-(B2-B29)-Y-X-Gly-(A2-A20)-R3 (II). In

CC formula (II), X = Lys, Arg or a peptide of 2-35 amino acids contg.

CC Lys or Arg at the N- and C-termini; Y = a natural amino acid; R1 =

CC Phe or a bond; R2 = H, Arg, Lys, a peptide of 2-45 amino acids

CC contg. Arg or Lys at the N- and C-termini; R3 = a natural amino

CC acid; (A2-A20) and (B2-B29) are the insulin A- and B-chain sequences

CC from human or other insulin. The proinsulin molecule, released by

CC cyanogen bromide, is reacted with mercaptan at a ratio of 2-10 SH

CC residues of mercaptan per Cys residue of proinsulin. The reaction

CC takes place in the presence of a chaotropic auxiliary agent at

CC pH 10-11 and results in proinsulin with correctly linked cystine

CC bridges. Reaction with trypsin and opt. carboxypeptidase B yields

CC correctly folded insulin. The insulin is isolated by absorption on

XX a hydrophobic resin.

XX Sequence 96 AA;

XX Query Match 50.9%; Score 299; DB 16; Length 96;

Best Local Similarity 100.0%; Pred. No. 5.le-18;  
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 55 RFVNHLCGSHLVEALVLCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107

DB 44 RFVNHLCGSHLVEALVLCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 96

RESULT 11

AAR71694

ID AAR71694 standard; protein: 115 AA.

XX AAR71694;

XX 25-MAR-2003 (updated)

XX 20-NOV-1995 (first entry)

XX Mating factor alpha 1-Insulin precursor ArgB1, ArgB31 N-terminal.

XX Human insulin precursor ArgB31; diabetes; Zinc ion complex;

XX mating factor alpha 1; N-terminal EEAEEAR.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Protein 1..85

FT /label= mating factor alpha-1

FT Peptide 86..93

FT /label= N-terminal peptide

FT Peptide 94..124

FT /label= B-chain

FT Peptide 125..145

FT /label= A-chain

XX W09507931-A1.

XX 23-MAR-1995.

XX 16-SEP-1994; 94WO-DK00347.

XX 17-SEP-1993; 93DK-00C1044.

XX 02-FEB-1994; 94US-0190829.

XX (NOVO) NOVO-NORDISK AS.

XX Andersen AS, Halstrom JB, Avelund S, Jonassen I;

XX Markussen J;

XX WPI: 1995-131314/17.

XX N-PSDB; AAQ86429.

XX Acylated insulin deriv. which may be present as a Zinc ion

XX complex - is used to treat diabetes and is rapid acting.

XX Example 5; Pages 82-83; 100pp; English.

XX AAQ86429 encodes AAR71694 mating factor alpha 1-Insulin precursor  
 CC ArgB1, ArgB31 N-terminal EEAEEAR. The insulin precursor comprises  
 CC the B and A chains of a claimed human insulin derivative preceded  
 CC by the N-terminal amino acids EEAEEAR. In the final claimed compsn.  
 CC they are covalently connected via disulphide bonds between Cys  
 CC residues A7/B7 and A20/B19. The derivative, which may be present  
 CC as a zinc ion complex, can be used as a fast action treatment for  
 CC diabetes.

XX (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 145 AA;

XX Query Match 50.9%; Score 299; DB 16; Length 145;

XX Best Local Similarity 100.0%; Pred. No. 7.3e-18;

XX Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 55 RFVNHLCGSHLVEALVLCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107

```
Db 93 RFVNOHLCGSHLVEALYLCGGERGFYTPKTRGIVEQCCTSIQSLYQLENYCN 145
|||||
RESULT 12
AAR71695
ID AAR71695 standard; Protein: 146 AA.
XX
AC AAR71695;
XX
DT 25-MAR-2003 (updated)
DT 20-NOV-1995 (first entry)
XX
XX Mating factor alpha 1-Insulin precursor ArgB1, ArgB31 N-terminal.
XX Human insulin precursor ArgB1, ArgB31; diabetes; zinc ion complex;
KW mating factor alpha 1; N-terminal EEAEAEAE.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Protein 1..85
FT /label= mating factor alpha-1
FT Peptide 86..94
FT /label= N-terminal peptide
FT Peptide 95..125
FT /label= B-chain
FT Peptide 126..146
FT /label= A-chain
XX
PN W09507931-Al.
XX
PD 23-MAR-1995.
XX
PF 16-SEP-1994; 94WO-DK00347.
XX
PR 17-SEP-1993; 93DK-0001044.
PR 02-FEB-1994; 94US-0190829.
XX
XX (NOVO ) NOVO-NORDISK AS.
XX
PI Andersen AS, Halstrom JB, Havelund S, Jonassen I;
PI Markussen J;
XX
DR WPI: 1995-131314/17.
DR N-PSDB: AAQ86432.
XX
PT Acylated insulin deriv. which may be present as a zinc ion
PT complex - is used to treat diabetes and is rapid acting.
XX
PS Example 6; Page 85; 100pp; English.
XX
CC AAQ86432 encodes AAR71695 mating factor alpha 1-Insulin precursor
CC ArgB1, ArgB31 N-terminal EEAEAEAE. The insulin precursor comprises
CC the B and A chains of a claimed human insulin derivative preceded
CC by the N-terminal amino acids EEAEAEAE. In the final claimed compn.
CC they are covalently connected via disulphide bonds between Cys
CC residues A7/B7 and A20/B19. The derivative, which may be present
CC as a zinc ion complex, can be used as a fast action treatment for
CC diabetes.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 146 AA;
Query Match 50.9%; Score 299; DB 16; Length 146;
Best Local Similarity 100.0%; Pred. No. 7.4e-18;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 55 RFVNOHLCGSHLVEALYLCGGERGFYTPKTRGIVEQCCTSIQSLYQLENYCN 107
|||||
Db 94 RFVNOHLCGSHLVEALYLCGGERGFYTPKTRGIVEQCCTSIQSLYQLENYCN 146
|||||
```

```
RESULT 13
AAY42859
ID AAY42859 standard; protein: 52 AA.
XX
AC AAY42859;
XX
DT 19-JAN-2000 (first entry)
XX
DE Human insulin precursor, SFO ID 5.
XX
KW Insulin; precursor; growth hormone; chaperone; intramolecular;
KW folding; conformation; chimeric protein; cleavable; recombinant;
KW production; yield.
XX
OS Homo sapiens.
XX
PN W09950302-Al.
XX
PD 07-OCT-1999.
XX
PF 31-MAR-1998; 98WO-CN00052.
XX
PR 31-MAR-1998; 98WO-CN00052.
XX
PA (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.
PI Gan Z;
XX
DR WPI: 1999-610839/52.
XX
XX New chimeric proteins containing human growth hormone fragment, used
PT particularly for the production of human insulin.
XX
PS Claim 12; Page 29-30; 46pp; English.
XX
CC This sequence represents a human insulin precursor comprising insulin
CC A and B chains. This insulin precursor is a component of the chimeric
CC proteins hGH-mini-proinsulin (AAY42860) and the chimeric protein
CC given in AAY42861. These chimeric proteins additionally contain an
CC N-terminal fragment of human growth hormone (hGH) and a cleavable
CC peptide linker (AAY42857). The hGH portion of the chimeric protein acts
CC as an intramolecular chaperone (IMC) for the insulin precursor,
CC enabling it to fold correctly. The cleavable peptide linker has a
CC C-terminal Arg residue which enables the hGH portion of the
CC chimeric protein to be removed after folding has taken place. Production
CC of recombinant human insulin via an hGH-proinsulin chimeric protein can
CC provide human insulin with correctly linked cysteine bridges with
CC fewer necessary procedural steps, and hence resulting in a higher yield
CC of human insulin. The IMC sequences not only protect insulin sequences
CC from intracellular degradation by a microorganism host, but also promote
CC the folding of the fused insulin precursor, facilitate the solubility of
CC the fusion protein and decrease the intermolecular interactions among
CC the fusion proteins, thus allowing folding of the fused insulin precursor
CC at commercially useful high concentrations. The procedural steps of
CC cyanogen bromide cleavage, oxidative sulphytolysis and related
CC purification steps can thus be eliminated, along with the use of high
CC concentrations of mercaptan or the use of hydrophobic absorbent resins.
XX
SQ Sequence 52 AA;
Query Match 50.1%; Score 294; DB 20; Length 52;
Best Local Similarity 100.0%; Pred. No. 7.8e-18;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 56 FVNOHLCGSHLVEALYLCGGERGFYTPKTRGIVEQCCTSIQSLYQLENYCN 107
|||||
Db 1 FVNOHLCGSHLVEALYLCGGERGFYTPKTRGIVEQCCTSIQSLYQLENYCN 52
|||||
RESULT 14
AAR04582
ID AAR04582 standard; protein: 57 AA.
XX
```

```

AC AAR04582;
XX
XX 25-MAR-2003 (updated)
DT 14-SEP-1990 (first entry)
XX
XX Proinsulin analogue with a Lys residue linking the A and B chains.
DE
XX insulin fusion protein; pro-insulin analogue; tendamistate;
KW Lys-Lys bridge; ds
XX
XX synthetic.
OS
XX
XX key Location/Qualifiers
FH misc-difference 36
FT /label=Lys residue linking insulin B chain to A chain
FT Peptide 1..35
FT /label= Insulin B chain
FT Peptide 37..57
FT /label= Insulin A chain
XX
XX EP367163-A.
PN
XX
XX 09-MAY-1990.
PD
XX
XX 28-OCT-19 9; 89EP-0120056.
PF
XX
XX 03-NOV-1988; 88DE-3837273.
PR
XX 19-AUG-1989; 89DE-3927449.
XX
XX (FARRH) HOECHST AG.
PA
XX
XX Koller KP, Riess GJ, Uhlmann E, Wallmeier H;
PI
XX
XX WPI; 1990-141149/19.
DR
XX N-PSDB; AAQ04335.
DR
XX
XX New insulin fusion proteins comprise pro-insulin analogue linked
PT to tendamistate
PT
XX
XX Disclosure; Page ?; 7pp; German.
PS
XX
XX This sequence is joined to the C-terminus of an N-terminal fragment
XX comprising opt. modified tendamistate. This fusion protein
CC may be converted into human insulin using known methods. The synthetic
CC gene was prepared by the phosphoramidite method.
CC
CC See also AAQ04336.
CC
CC (Updated on 25-MAR-2003 to correct PR field.)
CC
CC (Updated on 25-MAR-2003 to correct PI field.)
XX
XX Sequence 57 AA;
SQ
    Query Match 49.9%; Score 293; DB 11; Length 57;
    Best Local Similarity 96.2%; Pred. No. 1e-17; Indels 0; Gaps 0;
    Matches 51; Conservative 2; Mismatches 0;
QY 55 RFVNHQLGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107
    :|||||
DB 5 KFVNQLGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 57
    :|||||
RESULT 15
AAR79056
ID AAR79056 standard; Protein; 160 AA.
XX
XX AAR79056;
AC
XX
XX 25-MAR-2003 (updated)
DT 24-JAN-1996 (first entry)
XX
XX Glycosylphosphatidylinositol-anchored human recombinant insulin.
DE
XX
XX GPI; glycosylphosphatidylinositol; insulin; hormone; solubilization;
KW Saccharomyces cerevisiae; anchor; Gas1; plasmid pBY40.

```

```

XX Homo sapiens.
OS
XX
XX key Location/Qualifiers
FH Misc-difference 44..129
FT /note= "anchor attachment site"
XX
XX MO9522614-A1.
PN
XX 24-AUG-1995.
PD
XX
XX 16-FEB-1995; 95WO-BR00010.
PF
XX
XX 17-FEB-1994; 94BR-0000600.
PR
XX
XX (ESCO-) ESCOLA PAULISTA MEDICINA.
PA (FINE-) FINEP FINANCIADORA ESTUDOS & PROJETOS.
XX
XX Cardoso De Almeida ML, Anarai De Castilho Valavicius ;
PI Gomes De Amorim Filho A;
XX
XX WPI; 1995-302720/39.
DR
XX N-PSDB; AAQ99460.
DR
XX
XX Recombinant prodn. of proteins, e.g. insulin - by producing the
PT protein with a glycosylphosphatidyl:inositol anchor followed by
PT selective release
XX
XX Disclosure; Fig 3; 51pp; English.
PS
XX
XX Human recombinant insulin may be expressed in Saccharomyces
XX cerevisiae following linkage of the gene to the
CC glycosylphosphatidylinositol anchor. This anchoring technique can
CC provide for the release of the product in a highly specific and
CC selective manner. In addition, the recombinant protein will contain
CC an epitope which can be used in its final purification by
CC immunoaffinity. The protein product can be released by e.g. nitrous
CC deamination or treatment with neutral detergent.
CC (Updated on 25-MAR-2003 to correct PI field.)
XX
XX Sequence 160 AA;
SQ
    Query Match 49.1%; Score 288.5; DB 16; Length 160;
    Best Local Similarity 98.1%; Pred. No. 6.1e-17;
    Matches 53; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 55 RFVNHQLGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107
    :|||||
DB 43 RFVNHQLGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 96
    :|||||

```

Search completed: September 16, 2003, 12:38:26  
Job time : 33.642 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

protein - protein search, using sw model

on: September 16, 2003, 12:37:10 : Search time 17.07 Seconds  
(without alignments)  
265.217 Million cell updates/sec

file: US-09-423-100-6

reflect score: 587  
quence: 1 MFPTPLSLRFDNMLRAHR.....JVEQCCTSIICSLVLENYCN 107

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

sarched: 328717 seqs, 423108:8 residues

otal number of hits satisfying chosen parameters: 328717

imum DB seq length: 0  
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

1: /cgn2\_6/protdata/1/aaa/5A\_COMB.pep.\*  
2: /cgn2\_6/protdata/1/aaa/5B\_COMB.pep.\*  
3: /cgn2\_6/protdata/1/aaa/6A\_COMB.pep.\*  
4: /cgn2\_6/protdata/1/aaa/6B\_COMB.pep.\*  
5: /cgn2\_6/protdata/1/aaa/PTCUS\_COMB.pep.\*  
6: /cgn2\_6/protdata/1/aaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	Pos	ID	Description
1	304	51.8	63	1	US-08-160-376A-6	Sequence 6, Appl
2	302.5	51.5	137	1	US-08-400-256-39	Sequence 39, Appl
3	302.5	51.5	137	3	US-08-975-365-39	Sequence 39, Appl
4	299.5	51.0	66	1	US-08-291-0608-5	Sequence 5, Appl
5	299	50.9	56	1	US-08-160-376A-7	Sequence 7, Appl
6	299	50.9	56	1	US-08-389-487-11	Sequence 11, Appl
7	299	50.9	96	1	US-08-160-376A-5	Sequence 5, Appl
8	299	50.9	96	1	US-08-389-487-8	Sequence 8, Appl
9	299	50.9	145	1	US-08-400-256-45	Sequence 45, Appl
10	299	50.9	145	3	US-08-975-365-45	Sequence 45, Appl
11	299	50.9	146	1	US-08-400-256-48	Sequence 48, Appl
12	299	50.9	146	3	US-08-975-365-48	Sequence 48, Appl
13	293	49.9	57	1	US-08-030-731A-44	Sequence 44, Appl
14	287	48.9	65	3	US-08-900-574-3	Sequence 3, Appl
15	286.5	48.8	66	3	US-08-900-574-5	Sequence 5, Appl
16	286	48.7	67	3	US-08-900-574-7	Sequence 7, Appl
17	284.5	48.5	65	1	US-08-468-674B-71	Sequence 71, Appl
18	284.5	48.5	65	1	US-08-780-571-71	Sequence 71, Appl
19	284.5	48.5	124	3	US-09-012-669F-36	Sequence 36, Appl
20	284	48.4	138	3	US-08-932-082-19	Sequence 19, Appl
21	284	48.4	140	1	US-08-400-256-33	Sequence 33, Appl
22	284	48.4	140	1	US-08-400-256-42	Sequence 42, Appl
23	284	48.4	140	3	US-08-975-365-33	Sequence 33, Appl
24	284	48.4	140	3	US-08-975-365-42	Sequence 42, Appl
25	283.5	48.3	53	1	US-08-233-617-4	Sequence 4, Appl
26	283.5	48.3	53	4	US-08-981-988A-42	Sequence 42, Appl
27	283.5	48.3	117	3	US-09-012-669F-37	Sequence 37, Appl

28	281	47.9	104	1	US-08-400-256-15	Sequence 15, Appl
29	281	47.9	104	3	US-08-975-365-15	Sequence 15, Appl
30	280.5	47.8	89	1	US-08-468-674B-41	Sequence 41, Appl
31	280.5	47.8	89	1	US-08-780-571-41	Sequence 41, Appl
32	280.5	47.8	91	1	US-08-468-674B-45	Sequence 45, Appl
33	280.5	47.8	91	1	US-08-780-571-45	Sequence 45, Appl
34	280.5	47.8	124	1	US-08-446-646-3	Sequence 3, Appl
35	279.5	47.6	167	1	US-07-518-953-8	Sequence 8, Appl
36	279.5	47.6	167	1	US-08-081-661-8	Sequence 8, Appl
37	278.5	47.4	51	4	US-09-477-924-3	Sequence 3, Appl
38	278.5	47.4	51	4	US-09-723-981-3	Sequence 3, Appl
39	278.5	47.4	51	4	US-09-723-896-3	Sequence 3, Appl
40	278	47.4	117	4	US-09-280-030-63	Sequence 63, Appl
41	277.5	47.3	53	1	US-08-233-617-3	Sequence 1, Appl
42	277	47.2	96	2	US-09-134-836-4	Sequence 4, Appl
43	277	47.2	96	4	US-09-386-303A-4	Sequence 4, Appl
44	277	47.2	97	1	US-08-160-376A-4	Sequence 4, Appl
45	277	47.2	102	1	US-08-400-256-36	Sequence 36, Appl

#### ALIGNMENTS

#### RESULT 1

US-08-160-376A-6  
; Sequence 6, Application US/08160376A  
; Patent No. 5473049  
; GENERAL INFORMATION:  
; APPLICANT: Obermeier, Kanier  
; APPLICANT: Gerl, Martin  
; APPLICANT: Ludwig, Jurgen  
; APPLICANT: Sabel, Walter  
; TITLE OF INVENTION: Process For Obtaining Proinsulin  
; TITLE OF INVENTION: Possessing Correctly Linked  
; TITLE OF INVENTION: Cystine Bridges  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenneth A. Genoni, Esq.  
; STREET: Rt. 202-206 No. 5473049H/P.O. Box 2500  
; CITY: Somerville  
; STATE: New Jersey  
; COUNTRY: U.S.A.  
; ZIP: 08876-1258  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
; COMPUTER: IBM 386  
; OPERATING SYSTEM: WINDOWS 3.1  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08160376A  
; FILING DATE: December 1, 1993  
; CLASSIFICATION: 510  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GE P 4240420.7  
; FILING DATE: December 2, 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Barbara V. Maurer, Esq.  
; REGISTRATION NUMBER: 31,287  
; REFERENCE/DOCKET NUMBER: HOE 92/F 384  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (908) 231-4079  
; TELEFAX: (908) 231-2255  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 63 Amino Acids  
; TYPE: Amino Acid (AA)  
; TOPOLOGY: not relevant  
US-08-160-376A-6

Query Match 51.8% Score 304; DB 1; Length 63;  
Best Local Similarity 94.7% Pred No. 7.3e-29;  
Matches 54; Conservative 0; Mismatches 3; Indels 0; Gaps 0;



COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/291.0608  
FILING DATE: 08-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Einaudi, Carol P.  
REGISTRATION NUMBER: 32,220  
REFERENCE/DOCKET NUMBER: 02481.1105-02000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4366  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-291-0608-5

Query Match 51.0%; Score 299.5; DB 1; Length 66;  
Best Local Similarity 91.7%; Pred. No. 2.6e-28;  
Matches 55; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 48 NPLGTGPRVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107  
DB 8 DPNSNG-RFVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 66

RESULT 5  
US-08-160-376A-7  
Sequence 7, Application US/08160376A  
Patent No. 5473049  
GENERAL INFORMATION:  
APPLICANT: Obermeier, Ranier  
APPLICANT: Gerl, Martin  
APPLICANT: Ludwig, Jurgen  
APPLICANT: Sabel, Walter  
TITLE OF INVENTION: Process For Obtaining Proinsulin  
TITLE OF INVENTION: Possessing Correctly Linked  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenneth A. Genoni, Esq.  
STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500  
CITY: Somerville  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 08876-1258  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM 386  
OPERATING SYSTEM: WINDOWS 3.1  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/160.376A  
FILING DATE: December 1, 1993  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GE P 4240420.7  
FILING DATE: December 2, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Barbara V. Maurer, Esq.  
REGISTRATION NUMBER: 31,287  
REFERENCE/DOCKET NUMBER: HOE 92/F 384  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (908) 231-4079  
TELEFAX: (908) 231-2255  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 56 Amino Acids  
TYPE: Amino Acid (AA)  
TOPOLOGY: not relevant  
US-08-160-376A-7

Query Match 50.9%; Score 299; DB 1; Length 56;  
Best Local Similarity 100.0%; Pred. No. 2.4e-28;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 RFVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107  
DB 4 RFVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 56

RESULT 6  
US-08-389-487-11  
Sequence 11, Application US/08 389 487  
Patent No. 5663291  
GENERAL INFORMATION:  
APPLICANT: Obermeier, Ranier  
APPLICANT: Gerl, Martin  
APPLICANT: Ludwig, Jurgen  
APPLICANT: Sabel, Walter  
TITLE OF INVENTION: Process for Obtaining Insulin Having  
TITLE OF INVENTION: Correctly Linked Cystine Bridges  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/389.487  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Einaudi, Carol P.  
REGISTRATION NUMBER: 32,220  
REFERENCE/DOCKET NUMBER: 02481.1124-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 56 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-389-487-11

Query Match 50.9%; Score 299; DB 1; Length 56;  
Best Local Similarity 100.0%; Pred. No. 2.4e-28;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 RFVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107  
DB 4 RFVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 56

RESULT 7



US-08-160-376A-5  
: Sequence 5, Application US/08160376A  
: Patent No. 5473049  
: GENERAL INFORMATION:  
: APPLICANT: Obermeier, Rainer  
: APPLICANT: Gerl, Martin  
: APPLICANT: Ludwig, Jurgen  
: APPLICANT: Sabel, Walter  
: TITLE OF INVENTION: Process For Obtaining Proinsulin  
: TITLE OF INVENTION: Possessing Correctly Linked  
: TITLE OF INVENTION: Cystine Bridges  
: NUMBER OF SEQUENCES: 7  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Kenneth A. Genoni, Esq.  
: STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500  
: CITY: Somerville  
: STATE: New Jersey  
: COUNTRY: U.S.A.  
: ZIP: 08876-1258  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
: COMPUTER: IBM 386  
: OPERATING SYSTEM: WINDOWS 3.1  
: SOFTWARE: WORDPERFECT 5.1  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/160-376A  
: FILING DATE: December 1, 1993  
: CLASSIFICATION: 530  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: GE P 4240420.7  
: FILING DATE: December 2, 1992  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Barbara V. Maurer, Esq.  
: REGISTRATION NUMBER: 31,287  
: REFERENCE/DOCKET NUMBER: HOE 92/F 384  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (908) 231-4079  
: TELEFAX: (908) 231-2255  
: INFORMATION FOR SEQ ID NO: 5:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 96 Amino Acids  
: TYPE: Amino Acid (AA)  
: TOPOLOGY: not relevant  
US-08-160-376A-5

Query Match 50.9% Score 299; DB 1; Length 96;  
Best Local Similarity 100.0%; Pred. No. 4.8e-28;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107  
|||||  
Db 44 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 96

RESULT 8  
US-08-389-487-8  
: Sequence 8, Application US/08389487  
: Patent No. 5663291  
: GENERAL INFORMATION:  
: APPLICANT: Obermeier, Rainer  
: APPLICANT: Gerl, Martin  
: APPLICANT: Ludwig, Jurgen  
: APPLICANT: Sabel, Walter  
: TITLE OF INVENTION: Process for Obtaining Insulin Having  
: TITLE OF INVENTION: Correctly Linked Cystine Bridges  
: NUMBER OF SEQUENCES: 12  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
: DUNN  
: STREET: 1300 I Street, N.W.  
: CITY: Washington  
: STATE: D.C.  
: COUNTRY: United States of America

ZIP: 20005-3315  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/389,487  
: FILING DATE:  
: CLASSIFICATION: 530  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Einaudi, Carol P.  
: REGISTRATION NUMBER: 32,220  
: REFERENCE/DOCKET NUMBER: 02481.1424-00000  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 202-408-4000  
: TELEFAX: 202-408-4400  
: INFORMATION FOR SEQ ID NO: 8:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 96 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: peptide  
US-08-389-487-8

Query Match 50.9% Score 299; DB 1; Length 96;  
Best Local Similarity 100.0%; Pred. No. 4.8e-28;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107  
|||||  
Db 44 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 96

## RESULT 9

US-08-400-256-45  
: Sequence 45, Application US/08 00256  
: Patent No. 5750497  
: GENERAL INFORMATION:  
: APPLICANT: Havelund, Svend  
: APPLICANT: Halstrom, John  
: APPLICANT: Jonassen, Ib  
: APPLICANT: Andersen, Asger Sloth  
: APPLICANT: Markussen, Jan  
: TITLE OF INVENTION: ACYLATED INSULIN  
: NUMBER OF SEQUENCES: 49  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: No. 5750497o No. 5750497disk of No. 5750497th America, Inc.  
: STREET: 405 Lexington Avenue, 64th floor  
: CITY: New York  
: STATE: New York  
: COUNTRY: United States of America  
: ZIP: 10174-6401  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/400,256  
: FILING DATE: 03-MAR-1995  
: CLASSIFICATION: 514  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Landiris, Elias J.  
: REGISTRATION NUMBER: 33,728  
: REFERENCE/DOCKET NUMBER: 3985.220-US  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 212-867-0123  
: TELEFAX: 212-878-9655  
: INFORMATION FOR SEQ ID NO: 45:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 145 amino acids

```
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-400-256-45

Query Match
Best Local Similarity 100.0%; Score 299; DB 1; Length 145;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIICSLYLENYCN 107
Db 93 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIICSLYLENYCN 145

RESULT 10
US-08-975-365-45
; Sequence 45, Application US/08975365
; Patent No. 6011007
; GENERAL INFORMATION:
; APPLICANT: Havelund, Svend
; APPLICANT: Halstrom, John
; APPLICANT: Jonassen, Ib
; APPLICANT: Andersen, Asger Sloth
; APPLICANT: Markussen, Jan
; TITLE OF INVENTION: ACYLATED INSULIN
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5750497o No. 5750497disk of No. 6011007th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/400,256
; FILING DATE: 03-MAR-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3985.220-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 145 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-400-256-48

Query Match
Best Local Similarity 100.0%; Score 299; DB 1; Length 146;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIICSLYLENYCN 107
Db 94 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIICSLYLENYCN 146

RESULT 12
US-08-975-365-48
; Sequence 48, Application US/08975365
; Patent No. 6011007
; GENERAL INFORMATION:
; APPLICANT: Havelund, Svend
; APPLICANT: Halstrom, John
; APPLICANT: Jonassen, Ib
; APPLICANT: Andersen, Asger Sloth
; APPLICANT: Markussen, Jan
; TITLE OF INVENTION: ACYLATED INSULIN
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6011007o No. 6011007disk of No. 6011007th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/975,365  
FILING DATE: 05-MAY-1987  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 38 37 273.8  
FILING DATE: 03-NOV-1988  
APPLICATION NUMBER: US 08/400,256  
FILING DATE: 03-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 3985.220-US  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 48:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 146 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-975-365-48

Query Match 50.9%; Score 299; DB 3; Length 146;  
Best Local Similarity 100.0%; Pred. No. 8e-28;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 107  
DB 94 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 146

RESULT 13  
US-08-030-731A-44  
Sequence 44, Application US/08030731A  
Patent No. 5426036  
GENERAL INFORMATION:  
APPLICANT: Koller, Klaus-Peter  
APPLICANT: Riess, Guenther Johannes  
APPLICANT: Uhlmann, Eugen  
APPLICANT: Wallmeier, Holger  
TITLE OF INVENTION: Processes for the Preparation of Foreign  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/030,731A  
FILING DATE: 12-MAR-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/189,840  
FILING DATE: 03-MAY-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/430,622  
FILING DATE: 01-NOV-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/687,610  
FILING DATE: 19-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/735,757  
FILING DATE: 29-JUL-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 37 14 866.4

FILING DATE: 05-MAY-1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 38 37 273.8  
FILING DATE: 03-NOV-1988  
APPLICATION NUMBER: DE P 39 27 449.7  
FILING DATE: 19-AUG-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 40 12 818.0  
FILING DATE: 21-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Kirschner Michael K.  
REGISTRATION NUMBER: 34,851  
REFERENCE/DOCKET NUMBER: 02481-0593-02000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 44:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 57 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-030-731A-44

Query Match 49.9%; Score 293; DB 1; Length 57;  
Best Local Similarity 96.2%; Pred. No. 1.3e-27;  
Matches 51; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 55 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 107  
DB 5 RFVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 57

RESULT 14  
US-08-900-574-3  
Sequence 3, Application US/08900574  
Patent No. 6221837  
GENERAL INFORMATION:  
APPLICANT: Ertl, Johann  
APPLICANT: Habermann, Paul  
APPLICANT: Geisen, Karl  
APPLICANT: Seipke, Gerhard  
TITLE OF INVENTION: Insulin derivatives with increased zinc  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett,  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: District of Columbia  
COUNTRY: U.S.A.  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/900,574  
FILING DATE: July 24, 1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: German Application No. 6221837 19630242.0  
FILING DATE: July 26, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Carol P. Einaudi  
REGISTRATION NUMBER: 32,220  
REFERENCE/DOCKET NUMBER: 02481.1499-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 65 amino acids  
TYPE: Amino acid  
STRANDEDNESS: Single  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..65  
US-08-900-574-3

Query Match 48.9%; Score 287; DB 3; Length 65;  
Best Local Similarity 91.4%; Pred. No. 7.7e-27;  
Matches 53; Conservative 0; Mismatches 3; Indels 2; Gaps 1;  
QY 51 GTGPRFVNQHLGSGHLVEALYLVCGERGFFYTPKT--RGIVEQCCTSIICSLYQLENYC 106  
Db 7 GNSARFVNQHLGSGHLVEALYLVCGERGFFYTPKTHRHGIVEQCCTSIICSLYQLENYC 64

RESULT 15  
US-08-900-574-5  
Sequence 5, Application US/08900574  
Patent No. 6221837  
GENERAL INFORMATION:  
APPLICANT: Ertl, Johann  
APPLICANT: Habermann, Paul  
APPLICANT: Geisen, Karl  
APPLICANT: Seipke, Gerhard  
TITLE OF INVENTION: Insulin derivatives with increased zinc  
TITLE OF INVENTION: binding  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett,  
ADDRESSEE: & Dunner, L.L.P.  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: District of Columbia  
COUNTRY: U.S.A.  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/900,574  
FILING DATE: July 24, 1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: German Application No. 6221837 19630242.0  
FILING DATE: July 26, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Carol P. Elnaudi  
REGISTRATION NUMBER: 32,220  
REFERENCE/DOCKET NUMBER: 02481.1499-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 amino acids  
TYPE: Amino acid  
STRANDEDNESS: Single  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
ORIGINAL SOURCE:  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: Protein

LOCATION: 1..66  
US-08-900-574-5  
Query Match 48.8%; Score 286.5; DB 3; Length 66;  
Best Local Similarity 89.8%; Pred. No. 9e-27;  
Matches 53; Conservative 0; Mismatches 3; Indels 3; Gaps 1;  
QY 51 GTGPRFVNQHLGSGHLVEALYLVCGERGFFYTPKT--RGIVEQCCTSIICSLYQLENYC 106  
Db 7 GNSARFVNQHLGSGHLVEALYLVCGERGFFYTPKTAHHRGIVEQCCTSIICSLYQLENYC 65  
Search completed: September 16, 2003, 12:41:24  
Job time : 18.07 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 12:38:55 : Search time 33.7237 Seconds  
(without alignments)  
472.415 Million cell updates/sec

**Title:** us-09-423-100-6

**Perfect score:** 587

**Sequence:** 1 MEPTPLSLRFLDNLRAHR.....IVEQCCTSLSLYLENYCM 107

**Scoring table:** BLOSUM62

Gapop 10.0 , Gapext 0.5

**Searched:** 556269 seqs, 14893369 residues

**Total number of hits satisfying chosen parameters:** 556269

**Minimum DB seq length:** 0

**Maximum DB seq length:** 2000000000

**Post-processing:** Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

**Database:** 1: /cgn2\_6/ptodata/2/pubpa/US07\_PUBCOMB.pep.\*

2: /cgn2\_6/ptodata/2/pubpa/US07\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	587	100.0	107	14	US-10-054-873-6
2	555.5	94.6	150	14	US-10-054-873-7
3	294	50.1	152	14	US-10-054-873-5
4	284.5	48.5	124	10	US-09-894-711-18
5	284	48.4	138	10	US-09-861-687-19
6	278.5	47.4	51	11	US-09-858-935B-5
7	278.5	47.4	51	14	US-10-028-410-3
8	278	47.4	117	9	US-09-280-030-63
9	277	47.2	96	10	US-09-947-563-4
10	275.5	46.9	124	9	US-09-736-611-12
11	275.5	46.9	124	9	US-09-740-359-12
12	275.5	46.9	124	10	US-09-894-711-12
13	275.5	46.9	125	9	US-09-736-611-10
14	275.5	46.9	125	9	US-09-740-359-10
15	275.5	46.9	125	10	US-09-894-711-10

16 275.5 46.9 147 9 US-09-736-611-8  
17 275.5 46.9 147 9 US-09-740-359-7  
18 274 46.7 144 9 US-09-736-611-6  
19 274 46.7 144 9 US-09-740-359-5  
20 274 46.7 146 10 US-09-894-711-5  
21 273 46.5 50 14 US-10-066-009A-3  
22 271 46.2 96 10 US-09-947-563-5  
23 269.5 45.9 130 9 US-09-280-030-62  
24 267 45.5 86 11 US-09-878-380-1  
25 267 45.5 86 11 US-09-858-935B-4  
26 267 45.5 86 11 US-10-028-410-2  
27 267 45.5 86 14 US-10-054-873-4  
28 267 45.5 110 9 US-09-205-658-125  
29 267 45.5 110 9 US-09-815-229-3  
30 267 45.5 110 10 US-09-804-409A-9  
31 267 45.5 110 12 US-09-969-748C-6  
32 267 45.5 110 15 US-10-038-686-1  
33 267 45.5 110 15 US-10-328-813-2  
34 260 44.3 49 14 US-10-054-873-1  
35 260 44.3 92 14 US-10-054-873-2  
36 255.5 43.5 191 11 US-09-984-010-23  
37 255.5 43.5 191 12 US-10-153-207-1  
38 255.5 43.5 191 11 US-10-400-377-1  
39 255.5 43.5 191 11 US-10-400-708-1  
40 255.5 43.5 191 11 US-10-298-148-1  
41 255.5 43.5 214 11 US-10-153-207-6  
42 255.5 43.5 217 9 US-09-929-918-9  
43 255.5 43.5 245 9 US-09-280-030-66  
44 253.5 43.2 144 9 US-09-894-711-7  
45 249.5 42.5 217 9 S-09-853-688-2

#### ALIGNMENTS

#### RESULT 1

US-10-054-873-6  
: Sequence 6, Application US/1004873  
: Publication No. US20020164712A1  
: GENERAL INFORMATION:  
: APPLICANT: Gan, Zhong Ru  
: TITLE OF INVENTION: Chimeric Protein Containing an Intramolecular Chapterone-Like Sequence

NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/054,873

FILING DATE: 22-Jan-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/CN98/00052

FILING DATE: 31-MAR-1998

APPLICATION NUMBER: US 09/423,100

FILING DATE: 11-DEC-2000

ATTORNEY/AGENT INFORMATION:

NAME: Mycroft, Frank J

REGISTRATION NUMBER: 46,946

REFERENCE/DOCKET NUMBER: 020167-00013005

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 107 amino acids

TYPE: amino acid

Query Match	Best Local Similarity	Score	DB 14:	Length	DB 15:
Query Match	94.6%	Score 555.5;	DB 14:	Length 150;	
Best Local Similarity	71.3%	Pred. No. 5e-59;			
Matches 107;	Conservative	0;	Mismatches	0;	Indels 43; Gaps 1;
Db	1	MFPTPLSLRFDNAMLRAHRLHQLAFDYYQEFEEAYIPKEQKYSFLQNPLOTGTPRFFVNOH	60		
Db	1	MFPTPLSLRFDNAMLRAHRLHQLAFDYYQEFEEAYIPKEQKYSFLQNPLOTGTPRFFVNOH	60		
QY	61	LCGSHLVEALYLVCGRGFFYTPKTRIGIVEQVCTICSILYQLENYCN	107		
Db	61	LCGSHLVEALYLVCGRGFFYTPKTRIGIVEQVCTICSILYQLENYCN	107		
RESULT 2					
US-10-054-873-7					
Sequence 7;	Application US/10054873				
Publication No.	US20020164712A1				
GENERAL INFORMATION:					
APPLICANT:	Gan, Zhong Ku				
TITLE OF INVENTION:	Chimeric Protein Containing an Intramolecular Chaperone-Like Sequence				
NUMBER OF SEQUENCES:	7				
CORRESPONDENCE ADDRESS:					
ADDRESSEE:	Townsend and Townsend and Crew LLP				
STREET:	Two Embarcadero Center, Eighth Floor				
CITY:	San Francisco				
STATE:	California				
COUNTRY:	USA				
ZIP:	94111-3834				
COMPUTER READABLE FORM:					
MEDIUM TYPE:	Floppy disk				
COMPUTER:	IBM PC Compatible				
OPERATING SYSTEM:	PC-DOS/MS-DOS				
SOFTWARE:	PatentIn Release #1.0, Version #1.30				
CURRENT APPLICATION DATA:					
APPLICATION NUMBER:	US/10/054,873				
FILING DATE:	22-Jan-2002				
CLASSIFICATION:	<Unknown>				
PRIOR APPLICATION DATA:					
APPLICATION NUMBER:	WO PCT/CN98/000052				
FILING DATE:	31-MAR-1998				
APPLICATION NUMBER:	US 09/423,100				
FILING DATE:	11-DEC-2000				
ATTORNEY/AGENT INFORMATION:					
NAME:	Mycroft, Frank J				
REGISTRATION NUMBER:	46,946				
REFERENCE/DOCKET NUMBER:	020167-00013005				
INFORMATION FOR SEQ ID NO: 7:					
SEQUENCE CHARACTERISTICS:					
LENGTH:	150 amino acids				
TYPE:	amino acid				
STRANDEDNESS:	<Unknown>				
TOPOLOGY:	linear				
MOLECULE TYPE:	protein				
SEQUENCE DESCRIPTION:	SEQ ID NO: 7:				
US-10-054-873-7					
Query Match	94.6%	Score 555.5;	DB 14:	Length 150;	
Best Local Similarity	71.3%	Pred. No. 5e-59;			
Matches 107;	Conservative	0;	Mismatches	0;	Indels 43; Gaps 1;
Db	1	MFPTPLSLRFDNAMLRAHRLHQLAFDYYQEFEEAYIPKEQKYSFLQNPLOTGTPRFFVNOH	60		
Db	1	MFPTPLSLRFDNAMLRAHRLHQLAFDYYQEFEEAYIPKEQKYSFLQNPLOTGTPRFFVNOH	60		
QY	50	-----LGTGPRFVNOHLCGSHLVEALYLVCGR	77		

```

; PRIOR APPLICATION NUMBER: PA 2000 00443
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: PA 1999 01869
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/211,081
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/181,450
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: 09/740,359
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-894-711-18

Query Match      48.5%   Score 284.5; DB 10; Length 124;
Best Local Similarity 92.7%   Pred. No. 1.7e-26;
Matches 51; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY 54 PFVNOHLCGSHLVEALYLVCGERGFFYTPK-TRGIVEQCCTSCISLYOLENYCN 107
Db 70 PKFVNHLGCGSHLVEALYLVCGERGFFYTPKAAKGIVEQCCTSCISLYOLENYCN 124

RESULT 5
US-09-861-687-19
; Sequence 19, Application US/09861687
; Publication No. US20020193292A1
; GENERAL INFORMATION:
; APPLICANT: Markussen, Jan
; Jonassen, Ib
; Havelund, Svend
; Brandt, Jakob
; Kurtzhals, Peter
; Hansen, Hertz Per
; Kaarsholm, Niels Christian
; TITLE OF INVENTION: INSULIN DERIVATIVES
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. US20020193292A1o No. US /020193292A1disk of No. US200201932
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-Dos
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/861,687
; FILING DATE: 21-May-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/932,082
; FILING DATE: 16-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4341.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 138 amino acids
; TYPE: amino acid

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-861-687-19

Query Match      48.4%   Score 284; DB 10; Length 138;
Best Local Similarity 48.2%   Pred. No. 2.2e-26;
Matches 68; Conservative 5; Mismatches 28; Indels 40; Gaps 5;

QY 2 FPTIPLSRFLFDNMLRAHRLHQLAFDTYQFFEEAYTPKEQ--KVSFLO-----N 48
Db 3 FPSI-----FTAVLFAASSALAAPVNTTTEDEIAQIPAEAVIGYSNLEGFDFVAVLPFSN 57
QY 49 PLGTG-----PRFVNOHLCGSHLVEALYLVCGERGFFYTPK-TR 86
Db 58 STNNGLLFINTTIASTAAKEGVSLDKRFVNQHLGCGSHLVEALYLVCGERGFFYTPKAAK 117
QY 87 GIVEQCCTSCISLYOLENYCN 107
Db 118 GIVEQCCTSCISLYOLENYCN 138

RESULT 6
US-09-858-935B-5
; Sequence 5, Application US/09858935B
; Publication No. US20030069177F
; GENERAL INFORMATION:
; APPLICANT: Dubaquié, Yves
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Lowman, Henry B.
; TITLE OF INVENTION: METHOD FOR TREATING CARTILAGE DISORDERS
; FILE REFERENCE: P1794R1
; CURRENT APPLICATION NUMBER: US/09/858,935B
; CURRENT FILING DATE: 2002-07-02
; PRIOR APPLICATION NUMBER: US 60/248,985
; PRIOR FILING DATE: 2000-11-15
; PRIOR APPLICATION NUMBER: US 60/204,490
; PRIOR FILING DATE: 2000-05-16
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 1
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-858-935B-5

Query Match      47.4%   Score 278.5; DB 11; Length 51;
Best Local Similarity 98.1%   Pred. No. 2.9e-26;
Matches 51; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 107
Db 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 51

RESULT 7
US-10-028-410-3
; Sequence 3, Application US/10028410
; Publication No. US20020160955A1
; GENERAL INFORMATION:
; APPLICANT: Dubaquié, Yves
; APPLICANT: Lowman, Henry
; TITLE OF INVENTION: PROTEIN VARIANTS
; FILE REFERENCE: P1712R1-1
; CURRENT APPLICATION NUMBER: US/10/028,410
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US/09/477,924
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 3
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-028-410-3

```

```
Query Match          47.4%: Score 278.5; DB 14; Length 51;
Best Local Similarity 98.1%: Pred. No. 2.9e-26;
Matches 51; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 56 FVNHLCGSHLVEALYLCGERGFFYTPKTRGIVEQCCT-ICSLYOLENYCN 107
      |||||||
Db 1 FVNHLCGSHLVEALYLCGERGFFYTPKT-GIVEQCCT-ICSLYOLENYCN 51

RESULT 8
US-09-280-030-63
; Sequence 63, Application US/09280030A
; Patent No. US20010021515A1
; GENERAL INFORMATION:
; APPLICANT: Sato, Seiji
; APPLICANT: Higashikuni, Naohiko
; APPLICANT: Kudo, Toshiyuki
; APPLICANT: Kondo, Masaaki
; TITLE OF INVENTION: DNAS ENCODING NEW FUSION PROTEINS AND PROCESSES FOR
; TITLE OF INVENTION: PREPARING USEFUL POLYPEPTIDES THROUGH EXPRESSION OF THE
; FILE REFERENCE: 382.1026
; CURRENT APPLICATION NUMBER: US/09/280.030A
; EARLIER FILING DATE: 1999-03-26
; EARLIER FILING DATE: JP10-87339/1998
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Designated as
; OTHER INFORMATION: an amino acid sequence of
; OTHER INFORMATION: MWpmp10-Met-ProInsulin
US-09-280-030-63

Query Match          47.4%: Score 278; DB 9; Length 117;
Best Local Similarity 60.0%: Pred. No. 9.5e-26;
Matches 54; Conservative 0; Mismatches 2; Indels 34; Gaps 1;

Qy 52 TGPFRVNHLCGSHLVEALYLCGERGFFYTPKT----- 85
      |||||||
Db 28 TAPMFVNHLCGSHLVEALYLCGERGFFYTPKTRREADMLVGQVELGGPGAGSLQPL 87

Qy 86 -----RGIVEQCCT-ICSLYOLENYCN 107
      |||||||
Db 88 ALEGLSLQKRGIVEQCCT-ICSLYOLENYCN 117

RESULT 9
US-09-947-563-4
; Sequence 4, Application US/09947563
; Patent No. US20020156234A1
; GENERAL INFORMATION:
; APPLICANT: Rubroder, Franz-Josef
; APPLICANT: Keller, Reinhold
; TITLE OF INVENTION: Improved process for obtaining
; insulins precursors having correctly bonded cystine bridges
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farrahaw, Garrett &
; Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.10
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/947,563
FILING DATE: 07-Sep-2001
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 09/134,836
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Leslie McDonnell
REGISTRATION NUMBER: 34,872
REFERENCE/DOCKET NUMBER: 02481.1600-000000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 96 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: protein
LOCATION: 1..96
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-947-563-4

Query Match          47.2%: Score 277; DB 10; Length 96;
Best Local Similarity 59.3%: Pred. No. 9.8e-26;
Matches 54; Conservative 0; Mismatches 3; Indels 34; Gaps 1;

Qy 51 GTGPFRVNHLCGSHLVEALYLCGERGFFYTPKT----- 85
      |||||||
Db 6 GNSRFRVNHLCGSHLVEALYLCGERGFFYTPKTRREADMLVGQVELGGPGAGSLQPL 65

Qy 86 -----RGIVEC TSICSLYOLENYCN 107
      |||||||
Db 66 IALEGLSLQKRGIVEC TSICSLYOLENYCN 96

RESULT 10
US-09-736-611-12
; Sequence 12, Application US/09736611
; Patent No. US20010023069A1
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas
; APPLICANT: Ludvigsen, Svend
; APPLICANT: Kaarsholm, Niels
; TITLE OF INVENTION: Method For Making Insulin Precursors and
; TITLE OF INVENTION: Insulin Precursor Analogs
; FILE REFERENCE: 6058.200-US
; CURRENT APPLICATION NUMBER: US/09/736,611
; PRIOR FILING DATE: 2000-12-14
; PRIOR APPLICATION NUMBER: 60/181,443
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: 60/211,441
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: PA 1999 01868
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: PA 000 00440
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 124
; TYPE: PRT
; ORGANISM: N-terminal extension
US-09-736-611-12
```



```

; LENGTH: 711
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-894-711-12

Query Match          46.9%   Score 275.5; DB 10; Length 124;
Best Local Similarity 90.9%   Pred. No. 2.le-25;
Matches 50; Conservative      ; Mismatches 2; Indels 1; Gaps 1;

QY    54 PRFVNHLGCSHLEALYLVCGERGFYYTPKTRGVGEQCCTICSILYQLENYCN 107
Db     | :|:|||||||:|:|||||||:|:|||||||:|:|||||||:|:|||||||:|:|||||||
       70 PRFVNHLGCSHLEALYLVCGERGFYYTDKGKIVGEQCCTICSILYQLENYCN 124


RESULT 13
US-09-736-611-10
; Sequence 10, Application US/09736611
; Patent No. US20010023069A1
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas
; APPLICANT: Ludvigsen, Svend
; APPLICANT: Kaarshoim, Niels
; TITLE OF INVENTION: Method For Making Insulin Precursors and
; FILE REFERENCE: 6058.200-US
; CURRENT APPLICATION NUMBER: US/09736.611
; PRIOR FILING DATE: 2000-12-14
; PRIOR APPLICATION NUMBER: 60/181.443
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: 60/211.441
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: PA 1999 01868
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: PA 2000 00440
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 125
; TYPE: PRT
; ORGANISM: N-terminal extens.
US-09-736-611-10

Query Match          46.9%   Score 275.5; DB 9; Length 125;
Best Local Similarity 90.9%   Pred. No. 2.le-25;
Matches 50; Conservative      ; Mismatches 2; Indels 1; Gaps 1;

QY    54 PRFVNHLGCSHLEALYLVCGERGFYYTPKTRGVGEQCCTICSILYQLENYCN 107
Db     | :|:|||||||:|:|||||||:~|:|||||||:~|:|||||||:~|:|||||||:~|:|||||||
       71 PRFVNHLGCSHLEALYLVCGERGFYYTDKEWKGIVGEQCCTICSILYQLENYCN 125


RESULT 14
US-09-740-359-10
; Sequence 10, Application US/09740359
; Patent No. US20010041787A1
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas Borglum
; APPLICANT: Ludvigsen, Svend
; TITLE OF INVENTION: Method for making insulin precursors and
; TITLE OF INVENTION: insulin precursor analogues having improved fermentation
; TITLE OF INVENTION: yield in yeast
; FILE REFERENCE: 6148.200-US
; CURRENT APPLICATION NUMBER: US/09740.359
; CURRENT FILING DATE: 2000 3-19
; PRIOR APPLICATION NUMBER: PA 2000 00443
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: PA 1999 01869
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/211.081
; PRIOR FILING DATE: 2000-06-13
```

; PRIOR APPLICATION NUMBER: 60/181,450
; PRIOR FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 10
; LENGTH: 125
; TYPE: PRT
; ORGANISM: TA39 leader fused with N-terminally extended
US-09-740-359-10

Query Match 46.9% Score 275.5; DB 9; Length 125;
Best Local Similarity 90.9% Pred. No. 2.le-25;
Matches 50; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 54 PRFVNQHLGSHLVEALYLVCGERGFFYTPKT-RGIVEQCCTSIICSLYQLENYCN 107
I:|||||
DB 71 PKFVNQHLGSHLVEALYLVCGERGFFYTDKDGKGIIVEQCCTSIICSLYQLENYCN 125
I:|||||

RESULT 15
US-09-894-711-10
; Sequence 10, Application US/09894711
; Patent No. US20020137144A1
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas Borglum
; APPLICANT: Ludvigsen, Svend
; TITLE OF INVENTION: Method for making insulin precursors and
; TITLE OF INVENTION: insulin precursor analogues having improved fermentation
; TITLE OF INVENTION: yield in yeast
; FILE REFERENCE: 6148.400-US
; CURRENT APPLICATION NUMBER: US/09/894,711
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: PA 2000 00443
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: PA 1999 01869
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/211,081
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/181,450
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: 09/740,359
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-894-711-10

Query Match 46.9% Score 275.5; DB 10; Length 125;
Best Local Similarity 90.9% Pred. No. 2.le-25;
Matches 50; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 54 PRFVNQHLGSHLVEALYLVCGERGFFYTPKT-RGIVEQCCTSIICSLYQLENYCN 107
I:|||||
DB 71 PKFVNQHLGSHLVEALYLVCGERGFFYTDKDGKGIIVEQCCTSIICSLYQLENYCN 125
I:|||||

Search completed: September 16, 2003, 12:52:26
Job time : 34.7237 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 12:36:15 : Search time 12.4903 Seconds  
(without alignments)  
823.845 Million cell updates/sec

**Title:** US-09-423-100-6

**Perfect score:** 587

**Sequence:** 1 MFPTPLSLRFDNAMLRAHR.....IVEQCCTSIQSLYQLENYCN 107

**Scoring table:** BLOSUM62

Gapop 10.0 , Gapext 0.5

**Searched:** 283308 seqs, 96168682 residues

**Total number of hits satisfying chosen parameters:** 283308

**Minimum DB seq length:** 0

**Maximum DB seq length:** 2000000000

**Post-processing:** Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

**Database:** PIR76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	275	46.8	96	2	epidermal growth f
2	273.5	46.6	51	1	insulin - sperm wh
3	273.5	46.6	51	1	insulin - finback
4	273.5	46.6	51	1	insulin - elephant
5	273	46.5	110	2	insulin precursor
6	273	46.5	110	2	insulin precursor
7	271.5	46.3	51	1	insulin - hamster
8	268.5	45.7	51	1	insulin - Egyptian
9	267.5	45.6	51	2	insulin precursor
10	267	45.5	110	1	insulin precursor
11	267	45.5	110	2	insulin precursor
12	263.5	44.9	51	1	insulin - sei whal
13	263.5	44.9	51	1	insulin - goat
14	263.5	44.9	51	1	insulin - Arabian
15	263	44.8	84	1	insulin precursor
16	263	44.8	110	1	insulin precursor
17	262.5	44.7	51	1	insulin - cat
18	262	44.6	110	1	insulin precursor
19	261.5	44.5	51	1	insulin - common s
20	260	44.3	110	2	insulin precursor
21	258.5	44.0	105	1	insulin precursor
22	257	43.8	108	2	insulin precursor
23	256.5	43.7	51	2	insulin - North Am
24	255.5	43.5	217	1	insulin - North Am
25	255.5	43.5	217	2	somatotropin 1 pre
26	252.5	43.0	77	1	insulin precursor
27	252	42.9	86	1	insulin precursor
28	251.5	42.8	51	1	insulin - Chinchil
29	250	42.6	108	1	insulin 1 precursor

insulin 1 precursor  
insulin - goose  
insulin 2 precursor  
insulin 2 precursor  
insulin 11 - North  
insulin 12 - North  
insulin precursor  
insulin - crested  
insulin - turkey  
insulin - ostrich  
insulin - black-be  
insulin - slider t  
insulin precursor  
insulin - Amphiuma  
insulin - duckbill  
insulin precursor

#### ALIGNMENTS

##### RESULT 1

PC7082

epidermal growth factor/single chain insulin fusion protein - Bacillus brevis (fragme  
C:Species: Bacillus brevis  
C:Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 31-Mar-2003  
C:Accession: PC7082; PC7083

R:Koh, M.; Hanagata, H.; Ebisu, S.; Moribara, K.; Takagi, H.

Biosci. Biotechnol. Biochem. 64, 1079-1081, 2000

A:Title: Use of Bacillus brevis for synthesis and secretion of Des-B30 single-chain h

A:Reference number: PC7082; M010:20335834; PMID:10879487

A:Accession: PC7082

A:Molecule type: DNA

A:Residues: 1-96 <KOH>

A:Accession: PC7083

A:Molecule type: protein

A:Residues: 19-28 <K02>

C:Genetics:

A:Gene: egf-sci

C:Superfamily: insulin

Query Match

Best Local Similarity 46.8%; Score 275; DB 2; Length 96;

Matches 50; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

QY 55 RFVNOHLGSHLVEALYLVGGERGFYTPKTRGIVEQCCTSIQSLYQLENYCN 107

DB 46 RFVNOHLGSHLVEALYLVGGERGFYTPKTRGIVEQCCTSIQSLYQLENYCN 96

##### RESULT 2

INWHP

insulin - sperm whale

C:Species: Physeter catodon (sperm whale)

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jul-1999

C:Accession: A93142; A90082

R:Ishihara, Y.; Saito, T.; Ito, Y.; Fujino, M.

Nature 181, 1488-1469, 1958

A:Title: Structure of sperm- and sei-whale insulins and their breakdown by whale peps

A:Reference number: A93142

A:Accession: A93142

A:Molecule type: protein

A:Residues: 1-30; 31-51 <ISH>

R:Harris, J.I.; Sanger, F.; Naughton, M.A.

Arch. Biochem. Biophys. 65, 427-428, 1956

A:Title: Species differences in insulin.

A:Reference number: A90082

A:Accession: A90082

A:Molecule type: protein

A:Residues: 1-30; 31-51 <HAR>

C:Superfamily: insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status experimental <BCH>

F:1-30,31-51/Product: insulin #status experimental <MAT>  
 F:31-51/Domain: insulin chain A #status experimental <ACH>  
 F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 46.6%; Score 273.5; DB 1; Length 51;  
 Best Local Similarity 96.2%; Pred. No. 8.7e-22;  
 Matches 50; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107  
 |||||  
 Db 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCTSIQSLYLENYCN 51

## RESULT 3

INWHF

C:Species: Baleenoptera physalus (finback whale, common rorqual)  
 C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jul-1999  
 C:Accession: A91918

R:Hama, H.; Titani, K.; Sakaki, S.; Narita, K.

J. Biochem. 56, 285-293, 1964

A:Title: The amino acid sequence in fin-whale insulin.

A:Reference number: A91918

A:Accession: A5 918

A:Molecule type: protein

A:Residues: 1-30,31-51 &lt;HAM&gt;

C:Superfamily: insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status experimental &lt;BCH&gt;

F:30-31-51/Product: insulin #status experimental &lt;MAT&gt;

F:31-51/Domain: insulin chain A #status experimental &lt;ACH&gt;

F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 46.6%; Score 273.5; DB 1; Length 51;  
 Best Local Similarity 96.2%; Pred. No. 8.7e-22;  
 Matches 50; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107  
 |||||  
 Db 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCTSIQSLYLENYCN 51

## RESULT 4

INEL

C:Species: Elephantidae gen. sp. (elephant)  
 C:Date: 24-Apr-1984 #sequence\_revision 30-Sep-1988 #text\_change 16-Jul-1999  
 C:Accession: A01584

R:Smith, L.F.

Am. J. Med. 40, 662-666, 1966

A:Title: Species variation in the amino acid sequence of insulin.

A:Reference number: A90029; MUID:66160119; PMID:5949593

A:Accession: A01584

A:Molecule type: protein

A:Residues: 1-30,31-51 &lt;SMI&gt;

A:Note: the species of elephant is not given, but it is most probably the Indian elephant

C:Superfamily: insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status experimental &lt;BCH&gt;

F:30-31-51/Product: insulin #status experimental &lt;MAT&gt;

F:31-51/Domain: insulin chain A #status experimental &lt;ACH&gt;

F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 46.6%; Score 273.5; DB 1; Length 51;  
 Best Local Similarity 94.2%; Pred. No. 8.7e-22;  
 Matches 49; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

Qy 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYLENYCN 107  
 |||||  
 Db 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKT-GIVEQCCTGVCSLYLENYCN 51

## RESULT 5

## B42179

Insulin precursor - green monkey

C:Species: Cercopithecus aethiops (green monkey, grivet)

C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 16-Jul-1999

C:Accession: B42179; A05232; S16494; S22056

R:Seino, S.; Bell, G.I.; Li, W.H.

Mol. Biol. Evol. 9, 193-203, 1992

A:Title: Sequences of primates insulin genes support the hypothesis of a slower rate o

A:Reference number: A42179; MUID:92219953; PMID:1560757

A:Accession: B42179

A:Molecule type: DNA

A:Residues: 1-110 &lt;SEI&gt;

A:Cross-references: EMBL:X61092; NID:922808; PIDN:CAA43405.1; PID:g22809

A:Note: sequence extracted from NCBI backbone (NCBIN:95185, NCRIIP:95194)

R:Peterson, J.B.; Nehrlich, S.; Oyer, P.E.; Stelner, D.F.

J. Biol. Chem. 247, 4866-4871, 1972

A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog proin

A:Reference number: A92111; MUID:72258016; PMID:4626369

A:Accession: A05232

A:Molecule type: protein

A:Residues: 57-87 &lt;PET&gt;

C:Genetics:

A:Introns: 63/1

C:Superfamily: insulin

C:Keywords: hormone; pancreas

F:1-24/Domain: signal sequence #status predicted &lt;SIG&gt;

F:25-54/Domain: insulin chain B #status predicted &lt;BCH&gt;

F:57-87/Domain: connecting peptide #status predicted &lt;MAT&gt;

F:90-110/Domain: insulin chain A #status predicted &lt;ACH&gt;

F:31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 46.5%; Score 273; DB 2; Length 110;  
 Best Local Similarity 60.2%; Pred. No. 2.2e-21;  
 Matches 53; Conservative 0; Mismatches 1; Indels 34; Gaps 1;

Qy 54 PRFVNOHLCGSHLVEALYLVCGERGFFYTPKT----- 85

Db 23 PAFVNOHLCGSHLVEALYLVCGERGFFYTPKTRHAEADPOVGQVGLGGGRGAGSLQPLAL 82

Qy 86 -----RGIVEQCCTSIQSLYLENYCN 107

Db 83 EGSLOKRGIVEQCCTSIQSLYLENYCN 110

## RESULT 6

JQ0178

Insulin precursor - crab-eating macaque

C:Species: Macaca fascicularis (crab-eating macaque)

C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 16-Jul-1999

C:Accession: JQ0178

R:Wetkam, W.; Groneberg, J.; Leineweber, M.; Wengenmayer, F.; Winnacker, E.L.

Gene 19, 179-183, 1982

A:Title: The nucleotide sequence of cDNA coding for preproinsulin from the primate Ma

A:Reference number: JQ0178; MUID:83080474; PMID:6184262

A:Accession: JQ0178

A:Molecule type: mRNA

A:Residues: 1-110 &lt;WET&gt;

A:Cross-references: GB:J00336; NID:g342121; PIDN:AAA36849.1; PID:g342122

C:Superfamily: insulin

F:1-24/Domain: signal sequence #status predicted &lt;SIG&gt;

F:25-54,90-110/Product: insulin #status predicted &lt;MAT&gt;

F:55-89/Domain: insulin chain B #status predicted &lt;BCH&gt;

F:90-110/Domain: insulin chain A #status predicted &lt;ACH&gt;

F:31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 46.5%; Score 273; DB 2; Length 110;  
 Best Local Similarity 60.2%; Pred. No. 2.2e-21;  
 Matches 53; Conservative 0; Mismatches 1; Indels 34; Gaps 1;

Qy 54 PRFVNOHLCGSHLVEALYLVCGERGFFYTPKT----- 85

Db 23 PAFVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGVELGGPGAGSLQPLAL 82  
QY 86 -----RGIVEOCCTSIICSLYLENYCN 107  
Db 83 EGSLOKRGIVEOCCTSIICSLYLENYCN 110  
RESULT 7  
INHY  
insulin - hamster  
C:Species: Cricetinae gen. sp. (hamster)  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jul-1999  
C:Accession: A91456  
R:Neelon, F.A.; Delcher, H.K.; Steinman, H.; Lebovitz, H.E.  
Fed. Proc. 32, 300, 1973  
A:Title: Structure of hamster insulin: comparison with a tumor insulin.  
A:Reference number: A91456  
A:Accession: A91456  
A:Molecule type: protein  
A:Residues: 1-30; 31-51 <NEE>  
C:Superfamily: insulin  
C:Keywords: hormone; pancreas  
F:1-30/Domain: insulin chain B #status experimental <BCH>  
F:1-30;31-51/Product: insulin #status experimental <MAT>  
F:31-51/Domain: insulin chain A #status experimental <ACH>  
F:7-37;19-50;36-41/Disulfide bonds: #status predicted  
Query Match 46.3%; Score 271.5; DB 1; Length 51;  
Best Local Similarity 94.2%; Pred. No. 1.4e-21;  
Matches 49; Conservative 2; Mismatches 0; Indels 1; Gaps 1;  
QY 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEOCCTSIICSLYLENYCN 107  
Db 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKS-GIVDQCCTSIICSLYLENYCN 51  
RESULT 8  
INWSSP  
insulin - Egyptian splny mouse (tentative sequence)  
C:Species: Acomys cahirinus (Egyptian spiny mouse)  
C:Date: 13-Jul-1981 #sequence\_revision 13-Jul-1981 #text\_change 31-Mar-2000  
C:Accession: A01591  
R:Buenzli, H.F.; Humbel, R.E.  
Hoppe-Seyler's Z. Physiol. Chem. 353, 444-450, 1972  
A:Title: Isolation and partial structural analysis of insulin from mouse (Mus musculus)  
A:Reference number: A01591; MUID:72189454; PMID:5028210  
A:Contents: composition  
A:Accession: A01591  
A:Molecule type: protein  
A:Residues: 1-30; 31-51 <BUE>  
C:Superfamily: insulin  
C:Keywords: hormone; pancreas  
F:1-30/Domain: insulin chain B #status predicted <BCH>  
F:1-30;31-51/Product: insulin #status predicted <MAT>  
F:31-51/Domain: insulin chain A #status predicted <ACH>  
F:7-37;19-50;36-41/Disulfide bonds: #status predicted  
Query Match 45.7%; Score 268.5; DB 1; Length 51;  
Best Local Similarity 92.3%; Pred. No. 2.9e-21;  
Matches 48; Conservative 3; Mismatches 0; Indels 1; Gaps 1;  
QY 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEOCCTSIICSLYLENYCN 107  
Db 1 FVBOHLCGSHLVEALYLVCGERGFFYTPKS-GIVDQCCTSIICSLYLENYCN 51  
RESULT 9  
A59151  
insulin precursor - jack bean (fragments)  
N:Alternate names: hypoglycemic agent; plant  
C:Species: Canavalia ensiformis (jack bean)  
C:Date: 07-Dec-1999 #sequence\_revision 07-Dec-1999 #text\_change 10-Dec-1999  
C:Accession: B59151; A59151

R:Oliveira, A.E.A.; Machado, O.L.T.; Gomes, V.M.; Xavier-Neto, J.; Pereira, A.C.P.; V  
Protein Pept. Lett. 6, 15-21, 1999  
A:Title: Jack bean seed coat contains a protein with complete sequence homology to bo  
A:Reference number: A59151  
A:Accession: B59151  
A:Molecule type: protein  
A:Residues: 1-30 <MACB>  
A:Accession: A59151  
A:Molecule type: protein  
A:Residues: 31-51 <MACA>  
C:Comment: The two chains are probably produced from the same precursor:  
C:Superfamily: insulin  
F:1-30;31-51/Product: insulin #status experimental <MAT>  
F:1-30/Domain: chain B #status experimental <CHB>  
F:31-51/Domain: chain A #status experimental <CHA>  
F:7-37;19-50;36-41/Disulfide bonds: #status predicted  
Query Match 45.6%; Score 267.5; DB 2; Length 51;  
Best Local Similarity 92.3%; Pred. No. 3.7e-21;  
Matches 48; Conservative 1; Mismatches 2; Indels 1; Gaps 1;  
QY 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEOCCTSIICSLYLENYCN 107  
Db 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCASVCSLYLENYCN 51  
RESULT 10  
JPHU  
insulin precursor [validated] - human  
N:Alternate names: preproinsulin  
C:Species: Homo sapiens (man)  
C:Date: 23-Oct-1981 #sequence\_revision 23-Oct-1981 #text\_change 08-Dec 2000  
C:Accession: A93222; A94253; A93216; A93144; A92075; A91166; A91114; A01579;  
R:Bell, G.I.; Picot, R.L.; Rutter, W.J.; Cordell, B.; Tischer, E.; Goodman, H.M.  
Nature 284, 26-32, 1980  
A:Title: Sequence of the human insulin gene.  
A:Reference number: A93222; MUID:80120725; PMID:6243748  
A:Accession: A93222  
A:Molecule type: DNA  
A:Residues: 1-110 <REF>  
A:Cross-references: GB:J00265, D:q186429; PIDN:AAA59172.1; PID:q38682H  
R:Ullrich, A.; Dull, T.J.; Gray, A.; Brosius, J.; Sures, I.  
Science 209, 612-615, 1980  
A:Title: Genetic variation in the human insulin gene.  
A:Reference number: A94253; MUID:80236313; PMID:6248962  
A:Accession: A94253  
A:Molecule type: DNA  
A:Residues: 1-110 <ULL>  
A:Cross-references: GB:J00265; NID:q186429; PIDN:AAA59172.1; PID:q38682H  
R:Bell, G.I.; Swain, W.F.; Picot, R.; Cordell, B.; Goodman, H.M.; Rutter, W.J.  
Nature 282, 525-527, 1979  
A:Title: Nucleotide sequence of a cDNA clone encoding human preproinsulin.  
A:Reference number: A93216; MUID:80054779; PMID:503234  
A:Accession: A93216  
A:Molecule type: mRNA  
A:Residues: 1-110 <BEL2>  
A:Cross-references: GB:J00265; NID:q186429; PIDN:AAA59172.1; PID:q38682H  
R:Sures, I.; Goeddel, D.V.; Gray, A.; Ullrich, A.  
Science 208, 57-59, 1980  
A:Title: Nucleotide sequence of human preproinsulin complementary DNA.  
A:Reference number: A94251; MUID:80147417; PMID:6927840  
A:Accession: A94251  
A:Molecule type: mRNA  
A:Residues: 1-110 <SUR>  
A:Cross-references: GB:J00265; NID:q186429; PIDN:AAA59172.1; PID:q38682H  
R:Nicol, D.S.H.W.; Smith, L.F.  
Nature 187, 483-485, 1960  
A:Title: Amino-acid sequence of human insulin.  
A:Reference number: A93144  
A:Accession: A93144  
A:Molecule type: protein  
A:Residues: 25-54; 90-110 <NIC>  
R:Oyer, P.E.; Cho, S.; Peterson, J.D.; Steiner, D.F.

J. Biol. Chem. 246, 1375-1386, 1971  
A:Title: Studies on human proinsulin. Isolation and amino acid sequence of the human pan  
A:Reference number: A92075; MUID:71116410; PMID:5101771  
A:Accession: A92075  
A:Molecule type: protein  
A:Residues: 57-87 <OYE>  
R:Ko, A.; Smyth, D.G.; Markussen, J.; Sundby, F.  
Eur. J. Biochem. 20, 190-199, 1971  
A:Title: Amino acid sequence of the C-peptide of human proinsulin.  
A:Reference number: A91186; MUID:71257722; PMID:5560404  
A:Accession: A91186  
A:Molecule type: protein  
A:Residues: 57-87 <KOA>  
R:Lucassen, A.M.; Julier, C.; Beressi, J.P.; Boitard, C.; Froquel, P.; Lathrop, M.; Bell  
Nature Genet. 4, 305-310, 1993  
A:Title: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment  
A:Reference number: I58114; MUID:93364428; PMID:8358440  
A:Accession: I58114  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-59,63-110 <RES>  
A:Cross-references: GB:L15440; NID:q307071; PIDN:AAA59179.1; PID:q307072  
R:Sieber, P.; Kamber, B.; Hartmann, A.; Joehl, A.; Riniker, B.; Rittel, W.  
Helv. Chim. Acta 57, 2617-2621, 1974  
A:Title: Totalsynthese von Humaninsulin unter gezielter Bildung der Disulfidbindungen.  
A:Reference number: A91636; MUID:75077277; PMID:4443293  
A:Contents: annotation; synthesis  
A:Note: disulfide-bonded human insulin was synthesized; the synthetic hormone was identical  
R:Naithani, V.K.  
Hoppe-Seyler's Z. Physiol. Chem. 354, 659-672, 1973  
A:Title: The synthesis of C-peptide of human proinsulin.  
A:Reference number: A91658; MUID:75040007; PMID:4803504  
A:Contents: annotation; synthesis of residues 57-87  
R:Geiger, R.; Jaeger, G.; Koenig, W.  
Chem. Ber. 106, 2347-2352, 1973  
A:Title: Synthesis of the complete sequence of human proinsulin C-peptide and its [Glu-9  
A:Reference number: A90914  
A:Contents: annotation; synthesis of residues 57-87  
R:Kaufmann, J.E.; Irminger, J.C.; Halban, P.A.  
Biochem. J. 310, 869-874, 1995  
A:Title: Sequence requirements for proinsulin processing at the B-chain/C-peptide junction  
A:Reference number: S58661; MUID:96013185; PMID:7575420  
A:Contents: annotation; site-directed mutagenesis study of proteolytic processing  
C:Genetics:  
A:Gene: GDB:INS  
A:Cross-references: GDB:119349; OMIM:176730  
A:Map position: 11p15.5-11p15.5  
A:Introns: 63/1  
C:Superfamily: insulin  
C:Keywords: hormone; pancreas  
F:1-24/Domain: signal sequence \*status predicted <SIC>  
F:25-54/Domain: insulin chain B \*status experimental <BCH>  
F:55-110/Domain: insulin \*status experimental <MAT>  
F:57-87/Domain: connecting C peptide \*status experimental <CPEP>  
F:90-110/Domain: insulin chain A \*status experimental <ACH>  
F:31-96,43-109,95-100/Disulfide bonds: \*status experimental

Query Match 45.5% Score 267; DB 1; Length 110;  
Best Local Similarity 60.5% Pred. No. 9.3e-21;  
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;  
QY 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKT----- 85  
DB 25 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRAEADLQGVQLGGPGAGSLQPLALEG 84  
QY 86 ----RGIVEQCCTSCSLYLENYCN 107  
DB 85 SLOKRGIVEQCCTSCSLYLENYCN 110

RESULT 11  
A42179.

insulin precursor - chimpanzee  
C:Species: Pan troglodytes (chimpanzee)  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 16-Jul-1999  
C:Accession: A42179; S22058  
R:Seino, S.; Bell, G.I.; Li, W.H.  
Mol. Biol. Evol. 9, 193-203, 1992  
A:Title: Sequences of primate insulin genes support the hypothesis of a slower rate o  
A:Reference number: A42179; MUID:92219953; PMID:1560757  
A:Accession: A42179  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-110 <SEI>  
A:Cross-references: EMBL:X61089; NID:q38251; PIDN:CAA43403.1; PID:q38252  
A:Note: sequence extracted from NCBI backbone (NCBIP:95067)  
C:Genetics:  
A:Introns: 63/1  
C:Superfamily: insulin

Query Match 45.5% Score 267; DB 2; Length 110;  
Best Local Similarity 60.5% Pred. No. 9.3e-21;  
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;  
QY 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKT----- 85  
DB 25 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRAEADLQGVQLGGPGAGSLQPLALEG 84  
QY 86 ----RGIVEQCCTSCSLYLENYCN 107  
DB 85 SLOKRGIVEQCCTSCSLYLENYCN 110

RESULT 12  
INWHIS  
insulin - sei whale  
C:Species: Balaeoptera borealis (sei whale)  
C:Date: 13-Jul-1981 #sequence\_revision 13-Jul-1981 #text\_change 16-Jul-1999  
C:Accession: A01582  
R:Ishihara, Y.; Saito, T.; Ito, Y.; Fujino, M.  
Nature 181, 1468-1469, 1958  
A:Title: Structure of sperm- and sei-whale insulins and their breakdown by whale peps  
A:Reference number: A93142  
A:Accession: A01582  
A:Molecule type: protein  
A:Residues: 1-30,31-51 <ISH>  
C:Superfamily: insulin  
C:Keywords: hormone; pancreas  
F:1-30/Domain: insulin chain B \*status experimental <BCH>  
F:31-30,31-51/Product: insulin \*status experimental <MAT>  
F:31-51/Domain: insulin chain A \*status experimental <ACH>  
F:7-37,19-50,36-41/Disulfide bonds: \*status predicted

Query Match 44.9% Score 263.5; DB 1; Length 51;  
Best Local Similarity 92.3% Pred. No. 9.6e-21;  
Matches 48; Conservative 0; Mismatches 3; Indels 1; Gaps 1;  
QY 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCSLYLENYCN 107  
DB 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCTSCSLYLENYCN 51

RESULT 13  
INGT  
insulin - goat  
C:Species: Capra aegagrus hircus (domestic goat)  
C:Date: 13-Jul-1981 #sequence\_revision 13-Jul-1981 #text\_change 16-Jul-1999  
C:Accession: A01586  
R:Smith, L.F.  
Am. J. Med. 40, 662-666, 1966  
A:Title: Species variation in the amino acid sequence of insulin.  
A:Reference number: A90029; MUID:66160119; PMID:5949593  
A:Accession: A01586  
A:Molecule type: protein  
A:Residues: 1-30,31-51 <SMI>

C:Superfamily: insulin  
C:Keywords: hormone; pancreas  
F:1-30/Domain: insulin chain B #status experimental <BCH>  
F:1-30,31-51/Product: insulin #status experimental <MAT>  
F:31-51/Domain: insulin chain A #status experimental <ACH>  
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 44.9% Score 263.5; DB 1; Length 51;  
Best Local Similarity 90.4%; Pred. No. 9,6e-21;  
Matches 47; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

OY 56 FVNHLCGSHLVEALYLVCGERGFFYTPTKTRIGIVEQCCTSIICSLYLENYCN 107  
|||||  
DB 1 FVNHLCGSHLVEALYLVCGERGFFYTPTKA-GIVEQCCAGVCSLYLENYCN 51

RESULT 14  
INCA  
Insulin - Arabian camel (tentative sequence)  
C:Species: Camelus dromedarius (Arabian camel)  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jul-1999  
C:Accession: A92782  
R:banho, W.O.  
J. Fac. Med. Baghdad 14, 16-28, 1972  
A:Title: The isolation and characterization of insulin of camel (Camelus dromedarius).  
A:Reference number: A92782  
A:Accession: A92782  
A:Molecule type: protein  
A:Residues: 1-30;31-51 <DAN>  
C:Superfamily: insulin  
C:Keywords: hormone; pancreas  
F:1-30/Domain: insulin chain B #status experimental <BCH>  
F:1-30,31-51/Product: insulin #status experimental <MAT>  
F:31-51/Domain: insulin chain A #status experimental <ACH>  
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 44.9% Score 263.5; DB 1; Length 51;  
Best Local Similarity 90.4%; Pred. No. 9,6e-21;  
Matches 47; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

OY 56 FVNHLCGSHLVEALYLVCGERGFFYTPTKTRIGIVEQCCTSIICSLYLENYCN 107  
|||||  
DB 1 FVNHLCGSHLVEALYLVCGERGFFYTPTKA-GIVEQCCASVCSLYLENYCN 51

RESULT 15  
IPPG  
insulin precursor - pig  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 22-Jun-1981 #sequence\_revision 22-Jun-1981 #text\_change 16-Jul-1999  
C:Accession: A01583; A94572; S16492; A60835; B6(835  
R:Chance, R.E.; Ellis, R.M.; Bromer, W.W.  
Science 161, 165-167, 1968  
A:Title: Porcine proinsulin: characterization and amino acid sequence.  
A:Reference number: A94240; PMID:68286485; PMID:5657063  
A:Accession: A01583  
A:Molecule type: protein  
A:Residues: 1-34; Q'36-84 <CHA>  
R:Chance, R.E.  
submitted to the Atlas, July 1970  
A:Reference number: A94572  
A:Accession: A94572  
A:Molecule type: protein  
A:Residues: 1-84 <CH2>  
R:Brown, H.; Sanger, F.; Kital, R.  
Biochem. J. 60, 556-565, 1955  
A:Title: The structure of pig and sheep insulins.  
A:Reference number: A90344  
A:Accession: S16492  
A:Molecule type: protein  
A:Residues: 1-30;31-51 <BRO>  
R:Snel, L.; Damgaard, O.  
Horm. Metab. Res. 20, 476-480, 1988

A:Title: Proinsulin heterogeneity in pigs  
A:Reference number: A60835; PMID:9032178; PMID:3181865  
A:Accession: A60835  
A:Molecule type: protein  
A:Residues: 33-38,40-62 <SNE>  
A>Note: The authors report the characterization of a connecting peptide variant lacking  
A:Accession: B60835  
A:Molecule type: protein  
A:Residues: 33-62 <SN2>  
R:Blundell, T.; Dodson, G.; Hodakkin, D.; Mercola, D.  
Adv. Protein Chem. 26, 279-402, 1972  
A:Title: Insulin: the structure in the crystal and its reflection in chemistry and biology  
A:Reference number: A90017  
A:Contents: annotation: X-ray crystallography, 1.9 angstroms  
C:Superfamily: insulin  
C:Keywords: hormone; pancreas  
F:1-30/Domain: insulin chain B #status experimental <BCH>  
F:1-30,64-84/Product: insulin #status experimental <MAT>  
F:33-63/Domain: connecting peptide #status experimental <CPEP>  
F:64-84/Domain: insulin chain A #status experimental <ACH>  
F:7-70,19-83,69-74/Disulfide bonds: #status experimental

Query Match 44.8% Score 263; DB 1; Length 84;  
Best Local Similarity 60.7%; Pred. No. 1,8e-20;  
Matches 51; Conservative 0; Mismatches 1; Indels 32; Gaps 1;

OY 56 FVNHLCGSHLVEALYLVCGERGFFYTPTKTRARREAEFPQAGAVELGGIAGGLQAIALEGPP 60  
|||||  
DB 1 FVNHLCGSHLVEALYLVCGERGFFYTPTKARRREAENFQAGAVELGGIAGGLQAIALEGPP 60  
OY 86 --RGIVEQCCTSIICSLYLENYCN 107  
|||||  
DB 61 OKRGIVEQCCTSIICSLYLENYCN 84

Search completed: September 16, 2003, 12:40:35  
Job time : 12.4903 secs

GenCore version .1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 12:33:30 ; Search time 8.32685 Seconds  
(without alignments)  
604.293 Million cell updates/sec

**Title:** us-09-423-100-6

**Perfect score:** 587

**Sequence:** 1 MFPTIPLSRFLDNMLRAHR.....IVEQCCTICSGLYLENYCN 107

**Scoring table:** BLOSUM62

Gapop 10.0 , Gapext 0.5

**Searched:** 127863 seqs, 47026705 residues

**Total number of hits satisfying chosen parameters:** 127863

**Minimum DB seq length:** 0

**Maximum DB seq length:** 2000000000

**Post-processing:** Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

**Database :** SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	273.5	46.6	51	INS_BALPH	P01312 balaenopter
2	273.5	46.6	51	INS_ELMA	P01316 elephas max
3	273	46.5	110	INS_CERA	P30407 cercopthec
4	273	46.5	110	INS_MACFA	P30406 macaca fasc
5	268.5	45.7	51	INS_ACOCA	P01324 acomys cahi
6	267	45.5	110	INS_HUMAN	P01308 homo sapien
7	267	45.5	110	INS_PANTR	P30410 pan troglod
8	266	45.3	110	INS_SPETR	Q91x13 spermophilu
9	263.5	44.9	51	INS_BALBO	P01314 balaenopter
10	263.5	44.9	51	INS_CAMDR	P01320 camelus dro
11	263.5	44.9	51	INS_CAPHI	P01319 capra hircu
12	263	44.8	108	INS_PIG	P01315 sus scrofa
13	263	44.8	110	INS_RABIT	P01311 oryctolagus
14	262.5	44.7	51	INS_FELCA	P06306 felis silve
15	262	44.6	110	INS_CANFA	P01321 canis fami
16	260	44.3	110	INS_CRILLO	P01313 cricetus
17	258.5	44.0	105	INS_BOVIN	P01317 bos taurus
18	257	43.8	108	INS_AOTR	P10604 aotus trivi
19	257	43.8	110	INS_PSAOB	Q62587 psammomys o
20	256.5	43.7	51	INS_DIDMA	P18109 didelphis m
21	255.5	43.5	217	SOMA_HUMAN	P01241 homo sapien
22	255.5	43.5	217	SOMA_MACMU	P33093 macaca mula
23	255.5	43.5	217	SOMA_PANTR	P58756 pan troglod
24	254.5	43.4	105	INS_SHEEP	P01318 ovis aries
25	252	42.9	86	INS_HORSE	P01310 equus cabal
26	251.5	42.8	51	INS_CHTRB	P01327 chinchilla
27	250	42.6	108	INS_MOUSE	P01325 mus musculu
28	249	42.4	110	INSI_RAT	P01322 rattus norv
29	249	42.4	217	SOMA_CALJA	Q9gmb3 callithrix
30	249	42.4	217	SOMA_SAIBB	P58343 saimir bol
31	248.5	42.3	51	INS_ANGAN	P07454 anser anser
32	248	42.2	110	INS2_MOUSE	P01326 mus musculu
33	248	42.2	110	INS2_RAT	P01323 rattus norv

RESULT 1  
INS\_BALPH  
ID INS\_BALPH STANDARD: PRT: 51 AA.  
AC P01312:  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Insulin.  
GN INS.  
OS Balaenoptera physalus (Finback whale) (Common rorqual), and  
OS Physeter catodon (Sperm whale) (Physeter macrocephalus)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;  
OC Balaenopteridae; Balaenoptera.  
OX NCBI\_taxid=9770, 9755;  
RN [1]  
RP PARTIAL SEQUENCE.  
RC SPECIES-B.physalus;  
RA Hama H., Titani K., Sakaki S., Marita K.;  
RT "The amino acid sequence in fin-whale insulin.";  
RL J. Biochem. 56:285-293(1964)  
RN [2]  
RP SEQUENCE.  
RC SPECIES-P.catodon;  
RA Ishihara Y., Saito T., Ito Y., Fujino M.;  
RT "Structure of sperm- and sei-whale insulins and their breakdown by  
RT whale pepsin.";  
RL Nature 181:1468-1469(1958).  
RN [3]  
RP SEQUENCE.  
RC SPECIES-P.catodon;  
RA Harris J.I., Sanger F., Naughton M.A.;  
RT "Species differences in insulin.";  
RL Arch. Biochem. Biophys. 65:427-438(1956).  
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
CC DISULFIDE BONDS.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
DR PIR: A91918; INWHF.  
DR PIR: A93142; INWHP.  
DR HSSP: P01317; IAPH.  
DR InterP: IPR004825; Ins/IGF/relax.  
DR SMART: SM00078; IIGF; 1.  
DR PROSITE: PS00262; INSULIN; 1.  
KW Insulin family; Hormone; Glucose metabolism.  
FT CHAIN 1 30 INSULIN B CHAIN.  
FT NON-CONS 30 31  
FT CHAIN 31 51 INSULIN A CHAIN.  
FT DISULFID 7 37 INTERCHAIN.  
FT DISULFID 19 50 INTERCHAIN.  
FT DISULFID 36 41  
SQ SEQUENCE 51 AA: 5766 MW: 9007B514691A7CDD CRC64:

#### ALIGNMENTS

34 246 41.9 52 1 INS\_ACIGU  
35 245 41.7 103 1 INS\_SELRF  
36 244.5 41.7 51 1 INS\_HYSCR  
37 244.5 41.7 51 1 INS\_TRASC  
38 239.5 40.8 107 1 INS\_CHICK  
39 236 40.2 217 1 SOM2\_PANTR  
40 235.5 40.1 51 1 INS\_ORNAN  
41 233.5 39.8 81 1 INS\_ANAPL  
42 231.5 39.4 51 1 INS\_ALIMI  
43 231 39.4 52 1 INS\_LEPSP  
44 228.5 38.9 51 1 INS\_ZAODH  
45 228 38.8 217 1 SOM2\_HUMAN

p81423 acipenser g  
p51463 selasphorus  
p01328 hystrix cri  
p31887 trachemys s  
p01332 gallus gall  
p58757 pan troglod  
q9tqy7 ornithorhyn  
p01333 anas platyr  
p12703 alligator m  
p09476 lepisosteus  
p12708 zaccys dhum  
p01242 homo sapien



```

Query Match      46.6%; Score 273.5; DB 1; Length 51;
Best Local Similarity 96.2%; Pred. No. 1.6e-22;
Matches 50; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 56 FVNQHLGCSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 107
DB 1 FVNQHLGCSHLVEALYLVCGERGFFYTPKA-GIVEQCCTSCISLYOLENYCN 51

RESULT 2
INS_ELEMA
ID INS_ELEMA STANDARD; PRT: 51 AA.
AC P01316;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin.
GN INS.
OS Elephas maximus (Indian elephant).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Proboscidea; Elephantidae; Elephas.
OX NCBI_TaxID=9783;
RN [1]
RP SEQUENCE.
RX MEDLINE=66160119; PubMed=5949593;
RA Smith L.F.;
RL "Species variation in the amino acid sequence of insulin.";
CC "-1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC "-1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC "-1- SUBCELLULAR LOCATION: Secreted.
CC "-1- MISCELLANEOUS: THE SPECIES OF ELEPHANT IS NOT GIVEN, BUT IT IS
CC MOST PROBABLY THE INDIAN ELEPHANT (ELEPHAS MAXIMUS).
CC "-1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR HSSP: P01308; IAI0.
DR InterPro: IPR0004825; Ins/IGF/relax.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
FT SEQUENCE 51 AA: 5752 MW: 9007850CDB457D6D CR014;

Query Match      46.6%; Score 273.5; DB 1; Length 51;
Best Local Similarity 94.2%; Pred. No. 1.6e-22;
Matches 49; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

OY 56 FVNQHLGCSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSCISLYOLENYCN 107
DB 1 FVNQHLGCSHLVEALYLVCGERGFFYTPKT-GIVEQCCTGVCSLYOLENYCN 51

RESULT 3
INS_CERAE
ID INS_CERAE STANDARD; PRT: 110 AA.
AC P30407; P01309;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin precursor.
GN INS.
OS Cercopithecus aethiops (Green monkey) (Grivet).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;

```

```

Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=9534;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92219953; PubMed=1560757;
RA Seino S., Bell G.I., Li W.;
RT "Sequences of primate insulin genes support the hypothesis of a
RT slower rate of molecular evolution in humans and apes than in
RT monkeys.";
RL Mol. Biol. Evol. 9:193-203(1992).
RN [2]
RP SEQUENCE OF 57-87.
RX MEDLINE=72258016; PubMed=4626369;
RA Peterson J.D., Nehrlich S., Oyer P.E., Steiner D.F.;
RT "Determination of the amino acid sequence of the monkey, sheep, and
RT dog proinsulin C-peptides by a semi-micro Edman degradation
RT procedure.";
RL J. Biol. Chem. 247:4866-4871(1972).
CC "-1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC "-1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC "-1- SUBCELLULAR LOCATION: Secreted.
CC "-1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC EMBL: X61092; CAA43405.1;
DR PIR: B42179; B42179.
DR HSSP: P01308; IAI0.
DR InterPro: IPR0004825; Ins/IGF/relax.
DR Pfam: PF01049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 87 C-PEPTIDE.
FT CHAIN 90 110 INSULIN A CHAIN.
FT DISULFID 31 96 INTERCHAIN.
FT DISULFID 43 104 INTERCHAIN.
FT DISULFID 95 100
FT SEQUENCE 110 AA: 12019 MW: 9561548E7H247F4 CR014;

Query Match      46.5%; Score 273; DB 1; Length 110;
Best Local Similarity 60.2%; Pred. No. 4.1e-22;
Matches 53; Conservative 0; Mismatches 1; Indels 34; Gaps 1;

OY 54 PRFVNHLCGSHLVEALYLVCGERGFFYTPKT----- 85
DB 23 PAFVNHLCGSHLVEALYLVCGERGFFYTPKTREAEADPQVGVELGGGAGSLQPLAL 82

OY 86 -----RGIVEQCCTSCISLYOLENYCN 107
DB 83 EGSIQKRGIVEQCCTSCISLYOLENYCN 110

RESULT 4
INS_MACFA
ID INS_MACFA STANDARD; PRT: 110 AA.
AC P30406; P01309;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin precursor.

```

GN INS.  
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecoidea; Macaca.  
OX NCBI\_TaxID=9541;  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=63080474; PubMed=6184262;  
RA Metekam W., Groneberg J., Leineweber M., Wengenmayer F.,  
RA Winnacker E.-L.;  
RT "The nucleotide sequence of cDNA coding for preproinsulin from the  
RT primate Macaca fascicularis."  
RL Gene 19:179-183(1982).  
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
CC DISULFIDE BONDS.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: J00336; AAA36849.1;  
CC PIR: JQ0178; JQ0178.  
CC HSSP: P01308; LA10.  
DR InterPro: IPR004825; Ins/IGF/relax.  
DR Pfam: PF000049; Insulin: 1.  
DR PRINTS: PR00277; INSULINB.  
DR SMART: SM00078; IIGF: 1.  
DR PROSITE: PS00262; INSULIN: 1.  
KW Insulin family; Hormone; Glucose metabolism; Signal.  
FT SIGNAL 1 24  
FT CHAIN 25 54 INSULIN B CHAIN.  
FT PROPEP 57 87 C PEPTIDE.  
FT CHAIN 90 110 INSULIN A CHAIN.  
FT DISULFID 31 96 INSULIN A CHAIN.  
FT DISULFID 43 109 INTERCHAIN.  
FT DISULFID 95 100 INTERCHAIN.  
SQ SEQUENCE 110 AA; 11991 MW; 83C6E33A80A420F9 CRC64;

Query Match 46.5%; Score 273; DB 1; Length 110;  
Best Local Similarity 60.2%; Prod. No. 4.1e-22;  
Matches 53; Conservative 0; Mismatches 1; Indels 34; Gaps 1;  
  
Qy 54 PFVNHLCGSHLVEALYLVCGERGFFYTPKT----- 85  
Db 23 PFVNHLCGSHLVEALYLVCGERGFFYTPKTREAEADPQGVGLGGPGAGSLQPLAL 82  
  
Qy 86 -----RGIVEQCCTSCSLYLENYCN 107  
Db 83 EGSLOKRGIVEQCCTSCSLYLENYCN 110  
  
RESULT 5  
INS\_ACOCA  
ID INS\_ACOCA STANDARD; PRT; 51 AA.  
AC P01324;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Insulin.  
GN INS.  
OS Acomys cahirinus (Egyptian spiny mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Rodentia; Muridae; Murinae; Acomys.  
OX NCBI\_TaxID=10068;  
RN [1]  
RN COMPOSITION.  
RX MEDLINE=72189454; PubMed=5028210;  
RA Buentzli H.F., Humbel R.F.;  
RT "Isolation and partial structural analysis of insulin from mouse (Mus  
RT musculus) and spiny mouse (Acomys cahirinus)."  
RL Hoppe-Sevler's Z. Physiol. Chem. 353:444-450(1972).  
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
CC DISULFIDE BONDS.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
CC PIR: A01591; INMSSP.  
DR HSSP: P01308; IIGF: 1.  
DR InterPro: IPR004825; Ins/IGF/relax.  
DR SMART: SM00078; IIGF: 1.  
DR PROSITE: PS00262; INSULIN: 1.  
KW Insulin family; Hormone; Glucose metabolism.  
FT CHAIN 1 30 INSULIN B CHAIN.  
FT NON-CONS 30 31  
FT CHAIN 31 51 INSULIN A CHAIN.  
FT DISULFID 7 37 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 19 50 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 36 41 BY SIMILARITY.  
SQ SEQUENCE 51 AA; 5768 MW; 92B08B629047D3D CRC64;  
  
Query Match 45.7%; Score 268.5; DB 1; Length 51;  
Best Local Similarity 92.3%; Prod. No. 5.5e-22;  
Matches 48; Conservative 3; Mismatches 0; Indels 1; Gaps 1;  
  
Qy 56 PFVNHLCGSHLVEALYLVCGERGFFYTPKT----- 107  
Db 1 PFVNHLCGSHLVEALYLVCGERGFFYTPKS-GIVDQCCTSCSLYLENYCN 51  
  
RESULT 6  
INS\_HUMAN  
ID INS\_HUMAN STANDARD; PRT; 110 AA.  
AC P01308;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Insulin precursor.  
GN INS.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=80120725; PubMed=6243748;  
RA Bell G.I., Pictet R.L., Rutter W.J., Cordell B., Tischer E.,  
RA Goodman H.M.;  
RT "Sequence of the human insulin gene."  
RL Nature 284:26-32(1980).  
RN [2]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=80236313; PubMed=6248962;  
RA Ulrich A., Dull T.J., Gray A., Brosius J., Sures I.;  
RT "Genetic variation in the human insulin gene."  
RL Science 209:612-615(1980).  
RN [3]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=80054779; PubMed=503234;  
RA Bell G.I., Swain W.F., Pictet R.L., Cordell B., Goodman H.M.,  
RA Rutter W.J.;  
RT "Nucleotide sequence of a cDNA clone encoding human preproinsulin."  
RL Nature 282:525-527(1979).

- RN SEQUENCE FROM N.A.  
 RX MEDLINE-80147417; PubMed-6927840;  
 RA Sures I., Goeddel D.V., Gray A., Ullrich A.;  
 RT "Nucleotide sequence of human preproinsulin complementary DNA.";  
 RL Science 208:57-59(1980).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-93364428; PubMed-8358440;  
 RA Lucassen A.M., Bell J.I., Julier C., Lathrop M.;  
 RT "Susceptibility to insulin dependent diabetes mellitus maps to a 4.1  
 kb segment of DNA spanning the insulin gene and associated VNTR.";  
 RL Nat. Genet. 4:305-310(1993).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Pancreas;  
 RX MEDLINE-22388257; PubMed-12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.J., Wang J., Hsieh F.,  
 RA Datchenko L., Narusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [7]  
 RP SEQUENCE OF 1-59 FROM N.A.  
 RC TISSUE-Blood;  
 RX Pajardy I.I., Weill J.J., Stuckens C.C., Danze P.M.P.;  
 RA "Description of a novel RFLP diallelic polymorphism (-127 BsgI C/G)  
 within the 5' region of insulin gene.";  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP SEQUENCE OF 25-54 AND 90-110.  
 RA Nicol D.S.H.W., Smith L.F.;  
 RT "Amino-acid sequence of human insulin.";  
 RL Nature 187:483-485(1960).  
 RN [9]  
 RP SEQUENCE OF 57-87.  
 RX MEDLINE-71116410; PubMed-5101771;  
 RA Over P.E., Cho S., Peterson J.D., Steiner D.F.;  
 RT "Studies on human proinsulin. Isolation and amino acid sequence of  
 the human pancreatic C-peptide.";  
 RL J. Biol. Chem. 246:1375-1386(1971).  
 RN [10]  
 RP SEQUENCE OF 57-87.  
 RX MEDLINE-71257722; PubMed-5560404;  
 RA Ko A., Smyth D.G., Markussen J., Sundby F.;  
 RT "The amino acid sequence of the C-peptide of human proinsulin.";  
 RL Eur. J. Biochem. 20:190-199(1971).  
 RN [11]  
 RP SYNTHESIS.  
 RX MEDLINE-75077277; PubMed-4443293;  
 RA Steiber P., Kamber B., Hartmann A., Joehl A., Riniker B., Rittel W.;  
 RT "Total synthesis of human Insulin under directed formation of the  
 disulfide bonds.";  
 RL Helv. Chim. Acta 57:2617-2621(1974).  
 RN [12]  
 RP SYNTHESIS OF 57-87.  
 RX MEDLINE-75040007; PubMed-4803504;  
 RA Naithani V.K.;  
 RT "Studies on polypeptides, IV. The synthesis of C-peptide of human  
 proinsulin.";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 354:659-672(1973).  
 RN [13]  
 RP SYNTHESIS OF 65-69 AND 70-73.  
 RX MEDLINE-73161263; PubMed-4698555;  
 RA Geiger R., Volk A.;  
 RT "Synthesis of peptides with the properties of human proinsulin C  
 peptides (hC peptide). 3. Synthesis of the sequences 14-17 and 9-13  
 of human proinsulin C peptides.";  
 RL Chem. Ber. 106:199-205(1973).  
 RN [14]  
 RP SYNTHESIS OF 84-87.  
 RX MEDLINE-73161261; PubMed-4698553;  
 RA Geiger R., Jaeger G., Keonig W., Treuth G.;  
 RT "Synthesis of peptides with the properties of human proinsulin C  
 peptides (hC peptide). I. Scheme for the synthesis and preparation of  
 the sequence 28-31 of human proinsulin C peptide.";  
 RL Chem. Ber. 106:188-192(1973).  
 RN [15]  
 RP VARIANT LOS ANGELES SER-48.  
 RX MEDLINE-84016053; PubMed-6312455;  
 RA Haneda M., Chan S.J., Kwok S.C.M., Rubenstein A.H., Steiner D.F.;  
 RT "Studies on mutant human insulin genes: identification and sequence  
 analysis of a gene encoding [SerB24]insulin.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:6366-6370(1983).  
 RN [16]  
 RP VARIANTS LOS ANGELES SER-48 AND CHICAGO LEU-49.  
 RX MEDLINE-84170233; PubMed-6424111;  
 RA Shoelison S., Fickova M., Haneda M., Nahum A., Musso G., Kaiser E.T.,  
 RA Rubenstein A.H., Tager H.;  
 RT "Identification of a mutant human insulin predicted to contain a  
 serine-for-phenylalanine substitution.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:7390-7394(1983).  
 RN [17]  
 RP VARIANT PROVIDENCE ASP-34.  
 RX MEDLINE-87175640; PubMed-3477744;  
 RA Chan S.J., Seino S., Gruppas P.A., Schwartz R., Steiner D.F.;  
 RT "A mutation in the B chain coding region is associated with impaired  
 proinsulin conversion in a family with hyperproinsulinemia.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:2194-2197(1987).  
 RN [18]  
 RP VARIANT WAKAYAMA LEU-92.  
 RX MEDLINE-87058122; PubMed-3511;  
 RA Sakura H., Iwamoto Y., Sakuma Y., Kuzuya T., Hirata H.;  
 RT "Structurally abnormal insulin in a diabetic patient. Characterization  
 of the mutant insulin A3 (Val-->Leu) isolated from the pancreas.";  
 RL J. Clin. Invest. 78:1666-1672(1986).  
 RN [19]  
 RP VARIANT HIS-89.  
 RX MEDLINE-90317021; PubMed-2796279;  
 RA Barbetti F., Raben N., Kadowaki T., Cama A., Accilli D., Gabbay K.H.,  
 RA Merenich J.A., Taylor S.I., Roth J.;  
 RT "Two unrelated patients with familial hyperproinsulinemia due to a  
 mutation substituting histidine for arginine at position 65 in the  
 proinsulin molecule: identification of the mutation by direct  
 sequencing of genomic deoxyribonucleic acid amplified by polymerase  
 chain reaction.";  
 RL J. Clin. Endocrinol. Metab. 71:164-169(1990).  
 RN [20]  
 RP VARIANT HIS-89.  
 RX MEDLINE-85261996; PubMed-4019786;  
 RA Shibasaki Y., Kawakami T., Kanazawa Y., Akanuma Y., Takaku F.;  
 RT "Posttranslational cleavage of proinsulin is blocked by a point  
 mutation in familial hyperproinsulinemia.";  
 RL J. Clin. Invest. 76:378-383(1985).  
 RN [21]  
 RP VARIANT KYOTO LEU-89.  
 RX MEDLINE-92291307; PubMed-1601997;  
 RA Yano H., Kitano N., Morimoto M., Polonsky K.S., Imura H., Seino Y.;  
 RT "A novel point mutation in the human insulin gene giving rise to  
 hyperproinsulinemia (proinsulin Kyoto).";  
 RL J. Clin. Invest. 89:1902-1907(1992).

```

RN [22]
RP STRUCTURE BY NMR.
RX MEDLINE-91104966; PubMed-2271664;
RA Hua Q.-X., Weiss M.A.;
RT "Toward the solution structure of human insulin: sequential 2D 1H NMR
RT assignment of a des-pentapeptide analogue and comparison with crystal
RT structure.";
RL Biochemistry 29:10545-10555(1990).
RN [23]
RP STRUCTURE BY NMR.
RX MEDLINE-91242467; PubMed-2036420;
RA Hua Q.-X., Weiss M.A.;
RT "Comparative 2D NMR studies of human insulin and des-pentapeptide
RT insulin: sequential resonance assignment and implications for protein
RT dynamics and receptor recognition.";
RL Biochemistry 30:5505-5515(1991).
RN [24]
RP STRUCTURE BY NMR.
RX MEDLINE-91265527; PubMed-1646635;
RA Hua Q.-X., Weiss M.A.;
RT "Two-dimensional NMR studies of Des-(B26-B30)-insulin: sequence-
RT specific resonance assignments and effects of solvent composition.";
RL Biochim. Biophys. Acta 1078:101-110(1991).

Query Match 45.5%; Score 267; DB 1; Length 110;
Best Local Similarity 60.5%; Pred. No. 1.8e-21;
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKT----- 85
DB 25 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLVGQPGAGSLQPLALEG 84

QY 86 ----RGIVEQCCTSIQSLYLENYCN 107
DB 85 SLOKRGIVEQCCTSIQSLYLENYCN 110

RESULT 7
INS_PANTR
ID INS_PANTR STANDARD; PRT: 110 AA.
AC P30410;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Insulin precursor.
GN INS.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-9221953; PubMed=1560757;
RA Seleno S., Bell G.I., Li W.;
RT "Sequences of primate insulin genes support the hypothesis of a
RT slower rate of molecular evolution in humans and apes than in
RT monkeys.";
RL Mol. Biol. Evol. 9:193-203(1992).
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```

CC or send an email to license@isb-sib.ch.
CC EMBL; X61089; CAA43403.1; -.
DR PIR; A42179; A42179.
DR PDB; LEFE; 29-MAR-00.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN;
KW Insulin family; Hormone; Glucose metabolism; Signal; 3D-structure.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 87 C PEPTIDE.
FT CHAIN 90 110 INSULIN A CHAIN.
FT DISULFID 31 96 INTERCHAIN.
FT DISULFID 43 109 INTERCHAIN.
FT DISULFID 95 100
SQ SEQUENCE 110 AA; 12025 MW; 41EB8DF79837CEFS CRC64;

Query Match 45.5%; Score 267; DB 1; Length 110;
Best Local Similarity 60.5%; Pred. No. 1.8e-21;
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

QY 56 FVNOHLCGSHLVEALYLVCGERGFFYTPKT----- 85
DB 25 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQLVGQPGAGSLQPLALEG 84

QY 86 ----RGIVEQCCTSIQSLYLENYCN 107
DB 85 SLOKRGIVEQCCTSIQSLYLENYCN 110

RESULT 8
INS_SPETR
ID INS_SPETR STANDARD; PRT: 110 AA.
AC Q91X13;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Insulin precursor.
GN INS.
OS Spermophilus tridecemlineatus (Thirteen-lined ground squirrel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
OC Spermophilus.
OX NCBI_TaxID=43179;
RN [1]
RP SEQUENCE FROM N.A.
RX Tredrea M.M., Buck M.J., Gubanyiogi J., Squire T.L., Andrews M.T.;
RT "Regulation of PDK4 expression in a hibernating mammal.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch.
CC EMBL; AV038604; AAK72558.1; -.
DR HSSP; P01308; ILNP.
DR InterPro; IPR004825; Ins/IGF/relax.

```

```

DR Pfam: PF00049; Insulin: 1.
DR PRINTS: PR00277; INSULIN.
DR SMART: SM00078; IIGF: 1.
DR PROSITE: PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL 1 24 BY SIMILARITY.
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 87 C PEPTIDE.
FT CHAIN 90 110 INSULIN A CHAIN.
FT DISULFID 31 96 INTERCHAIN (BY SIMILARITY).
FT DISULFID 43 109 INTERCHAIN (BY SIMILARITY).
FT DISULFID 95 100 BY SIMILARITY.
SQ SEQUENCE 110 AA; 12004 MW; 4511768D6122BEE5 CRC64;

Query Match 45.3%; Score 266; DB 1; Length 110;
Best Local Similarity 57.4%; Pred. No. 2.2e-21;
Matches 54; Conservative 1; Mismatches 3; Indels 36; Gaps 2;

QY 50 LGTGP--RFVNHLCGSHLVYLVCGERGFYTPKTRGIVEOCCTSIICSLYLENYCN 107
DB 17 LGPDPAQAFVNHLCGSHLVYLVCGERGFYTPKSRREVVEEQGGQVFLGGPGAGL 76
QY 85 -----RGIVEOCCTSIICSLYLENYCN 107
DB 77 PQPLALEMALQKRGIVEOCCTSIICSLYLENYCN 110

RESULT 9
INS_BALBO
ID INS_BALBO STANDARD; PRT: 51 AA.
AC P01314;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Insulin.
GN INS.
OS Balaenoptera borealis (Sei whale).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
OC Balaenopteridae; Balaenoptera.
OX NCBI_TaxID=9768;
RN [1]
RP SEQUENCE.
RA Ishihara Y., Saito T., Ito Y., Fujino M.;
RT "Structure of sperm- and sei-whale insulins and their breakdown by
RT whale pepsin."
RL Nature 181:1468-1469(1958).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: A01582; INWHIS.
DR HSSP: P01317; IAPH.
DR SMART: PR004825; Ins/IGF/relax.
DR PROSITE: SM00078; IIGF: 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5723 MW; 9007B50E400A7DDD CRC64;

Query Match 44.9%; Score 263.5; DB 1; Length 51;
Best Local Similarity 92.3%; Pred. No. 1.9e-21;
Matches 48; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 56 FVNHLCGSHLVYLVCGERGFYTPKTRGIVEOCCTSIICSLYLENYCN 107
DB 1 FVNHLCGSHLVYLVCGERGFYTPKTRGIVEOCCTSIICSLYLENYCN 51

RESULT 10
INS_CAMDR
ID INS_CAMDR STANDARD; PRT: 51 AA.
AC P01320;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin.
GN INS.
OS Camelus dromedarius (Dromedary) (Arabian camel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OX NCBI_TaxID=9838;
RN [1]
RP SEQUENCE.
RA Danho W.O.;
RT "The isolation and characterization of insulin of camel (Camelus
RT dromedarius)."
RL J. Fac. Med. Baghdad 14:16-28(1972).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: A92782; INCOMA.
DR HSSP: P01317; 2INS.
DR SMART: SM00078; IIGF: 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5693 MW; 901E88BA085A7DDD CRC64;

Query Match 44.9%; Score 261.5; DB 1; Length 51;
Best Local Similarity 90.4%; Pred. No. 1.9e-21;
Matches 47; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 56 FVNHLCGSHLVYLVCGERGFYTPKTRGIVEOCCTSIICSLYLENYCN 107
DB 1 FVNHLCGSHLVYLVCGERGFYTPKTRGIVEOCCTSIICSLYLENYCN 51

RESULT 11
INS_CAPHI
ID INS_CAPHI STANDARD; PRT: 51 AA.
AC P01319;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin.
GN INS.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE.
RA MEDLINE=66160119; PubMed=5949593;
RA Smith L.F.;

```

```

RT *Species variation in the amino acid sequence of insulin.*
RL Am. J. Med. 40:662-666(1966).
CC -I- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. I1
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -I- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: A01586; INGT.
DR HSSP: P01317; IAPH.
DR InterPro: IPR004825; Ins/IGF/relax.
DR SMART: SM00078; ILGF: 1.
DR PROSITE: PS00262; INSULIN: 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON-CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA: 5692 MW: 9007850CD84E7DDD CRC64:
Query Match 44.9% Score 263.5; DB 1; Length 51;
Best Local Similarity 90.4%; Pred. No. 1.9e-21;
Matches 47; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
QY 56 FVNHLCGSHLVEALYLVCGERGFFYTPKTRIGVEQCCTISCSLYOLENYCN 107
DDB 1 FVNHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCAGVCSLYOLENYCN 5;
RESULT 12
INS_PIG
ID INS_PIG STANDARD: PRT: 108 AA.
AC P01315; OQTSJ5;
DI 21-JUL-1986 (Rel. 01, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Insulin precursor.
GN INS.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Han X.G., Tsch B.E.;
RT Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Large white;
RX MEDLINE=22135958; PubMed=12140686;
RA Amarger V., Nguyen M., Laere A.S., Braunschweig M., Nezer C.,
RA Georges M., Andersson L.;
RT Comparative sequence analysis of the INS-IGF2-H19 gene cluster in
RT pigs.*
RL Mamm. Genome 13:388-398(2002).
RN [3]
RP SEQUENCE OF 25-108.
RX MEDLINE=68286485; PubMed=5657063;
RA Chance R.E., Ellis R.M., Bromer W.W.;
RT *Porcine proinsulin: characterization and amino acid sequence.*;
RL Science 161:165-167(1968).
RN [4]
RP REVISION TO 59.
RA Chance R.E.;
RL Submitted (JUL-1970) to the PIR data bank.
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RA Blundell T.L., Dodson G.G., Hodgkin D., Mercola D.;

```

```

RT *Insulin. The structure in the crystal and its reflection in
RT chemistry and biology.*
RL Adv. Protein Chem. 26:279-402(1972).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RA Isaacs N.W., Agarwal R.C.;
RT *Experience with fast Fourier least squares in the refinement of the
RT crystal structure of rhombohedral 2-zinc insulin at 1.5-A
RT resolution.*;
RL Acta Crystallogr. A 34:782-791(1978).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RX MEDLINE=89099418; PubMed=2905485;
RA Baker E.N., Blundell T.L., Cutfield J.F., Cutfield S.M., Dodson E.J.,
RA Dodson G.G., Crowfoot Hodgkin D.M., Hubbard R.E., Isaacs N.W.,
RA Reynolds C.D., Sakabe K., Sakabe N., Vijayan N.M.;
RT *The structure of 2Zn pig insulin crystals at 1.5-A resolution.*;
RL Philos. Trans. R. Soc. Lond. B, Biol. Sci. 319:369-456(1988).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=92126280; PubMed=1772633;
RA Balschmidt P., Hansen F.B., Dodson E., Dodson G., Kothe F.;
RT *Structure of porcine insulin co-crystallized with clupeine Z.*;
RL Acta Crystallogr. B 47:975-986(1991).
RN [9]
RP X-RAY CRYSTALLOGRAPHY.
RX MEDLINE=91222450; PubMed=2025410;
RA Badger J., Harris M.R., Reynolds G.D., Evans A.C., Dodson E.J.,
RA Dodson G.G., North A.C.J.;
RT *Structure of the pig insulin dimer in the cubic crystal.*;
RL Acta Crystallogr. B 47:127-136(1991).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).
RA Diau J.-S., Wan Z.-L., Chang W.-R., Liang D.-C.;
RT *Structure of monomeric porcine DesB1-B2 desptapeptide (B26-B46)
RT insulin at 1.65-A resolution.*;
RL Acta Crystallogr. D 53:507-512(1997).
CC -I- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -I- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -I- DATABASE: NAME-Protein Spotlight;
CC NOTE=Issue 9 of April 2001;
CC WWW="http://www.expasy.org/spotlight/articles/spt1009.html".
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: AF064555; AAC77920.1; ALT_INIT.
DR EMBL: AY044828; AAL69550.1;
DR PDB: 3INS; 09-JAN-89.
DR PDB: 4INS; 31-JUL-94.
DR PDB: 6INS; 31-JAN-94.
DR PDB: 7INS; 31-JAN-94.
DR PDB: 9INS; 15-OCT-91.
DR PDB: 11ZA; 15-OCT-91.
DR PDB: 11ZB; 15-OCT-91.
DR PDB: 12CI; 29-JAN-96.
DR PDB: 1MPJ; 29-JAN-96.
DR PDB: 3MTH; 29-JAN-96.
DR PDB: 1DEI; 16-JUN-97.
DR PDB: 1SDB; 01-APR-98.
DR PDB: 1WAV; 28-FEB-97.
DR PDB: 1ZEI; 16-FEB-99.

```

```

DR PDB; 1ZNI; 28-JAN-98.
DR PDB; 1ZNI; 28-JAN-98.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal; 3D-structure.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 85 C PEPTIDE.
FT CHAIN 88 108 INSULIN A CHAIN.
FT DISULFID 31 94 INTERCHAIN.
FT DISULFID 43 107 INTERCHAIN.
FT DISULFID 93 98
FT HELIX 26 46
FT STRAND 48 48
FT STRAND 89 94
FT HELIX 100 106
FT STRAND 107 107
SQ SEQUENCE 108 AA; 11671 MW; CB491B429858EBC CRC64;

Query Match 44.8%; Score 263; DB 1; Length 108;
Best Local Similarity 60.7%; Pred. No. 4.6e-21;
Matches 51; Conservative 0; Mismatches 1; Indels 32; Gaps 1;

QY 56 FVNHLCGSHLVEALYLVCGERGFFYTPKT----- 85
DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKRRARENPAQAVELGGGLGQLALAEQGP 84

QY 86 --RGIVEQCCTSCSYLYENYCN 107
DB 85 QKRGIVEQCCTSCSYLYENYCN 108

RESULT 13
ID INS_RAB11 STANDARD; PRT; 110 AA.
AC P01311;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Insulin precursor.
GN INS.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-New Zealand white; ISSUE-Pancreas;
RX MEDLINE=94179230; PubMed=8132571;
RA Devaskar S.U.; Giddings S.J.; Rajakumar P.A.; Carnaghi L.R.;
RA Menon R.K.; Zahm D.S.;
RT "Insulin gene expression and insulin synthesis in mammalian neuronal
cells.";
RN J. Biol. Chem. 269:8445-8454(1994).
RL [2]
RP SEQUENCE OF 25-54 AND 90-110.
RX MEDLINE=66160119; PubMed=594593;
RA Smith L.F.;
RT "Species variation in the amino acid sequence of insulin.";
RL Am. J. Med. 40:662-666(1966).
RN [3]
RP SEQUENCE OF 56-110 FROM N.A.
RX Giddings S.J.; Carnaghi L.R.; Devaskar S.U.;
RA Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: A01588; INCT.
DR HSSP: P01317; IAPH.
DR InterPro: IPR004825; Ins/IGF/relax.
DR PRINTS; PR00277; INSULINB.

```

CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U03610; AAA19033.1; .

CC EMBL; M61153; AAA17540.1; .

DR PIR: A53438; INRB.

DR HSSP: P0108; ILYM.

DR Pfam: PF00049; Insulin; 1.

DR InterPro: IPR004825; Ins/IGF/relax.

DR SMART: SM00078; IIGF; 1.

DR PROSITE: PS00262; INSULIN; 1.

KW Insulin family; Hormone; Glucose metabolism; Signal.

FT SIGNAL 1 24

FT CHAIN 25 54 INSULIN B CHAIN.

FT PROPEP 57 87 C PEPTIDE.

FT CHAIN 90 110 INSULIN A CHAIN.

FT DISULFID 31 96 INTERCHAIN.

FT DISULFID 43 109 INTERCHAIN.

FT DISULFID 95 100

FT CONFLICT 83 83 E -> Y (IN REF. 3).

SQ SEQUENCE 110 AA; 11638 MW; 82D2975865D17FA8 CRC64;

Query Match 44.8%; Score 263; DB 1; Length 110;

Best Local Similarity 59.3%; Pred. No. 4.6e-21;

Matches 51; Conservative 1; Mismatches 0; Indels 34; Gaps 1;

QY 56 FVNHLCGSHLVEALYLVCGERGFFYTPKT----- 85

DB 25 FVNHLCGSHLVEALYLVCGERGFFYTPKSRREVELVGVQAEIIGCGKAGLQPSALEL 84

QY 86 ----RGIVEQCCTSCSYLYENYCN 107

DB 85 ALQKRGIVEQCCTSCSYLYENYCN 110

RESULT 14

ID INS\_FELCA STANDARD; PRT; 51 AA.

AC P00306;

DT 01-JAN-1988 (Rel. 06, Created)

DT 01-JAN-1988 (Rel. 06, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Insulin.

GN INS.

OS Felis silvestris catus (Cat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.

OX NCBI\_TaxID=9685;

RN [1]

RP SEQUENCE.

RX MEDLINE=86214076; PubMed=3518635;

RA Halliden G.; Gavellin G.; Mutt V.; Joernvall H.;

RT "Characterization of cat insulin."

RL Arch. Biochem. Biophys. 247:20-27(1986).

CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT

CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND

CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE

CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.

CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO

CC DISULFIDE BONDS.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

DR PIR: A01588; INCT.

DR HSSP: P01317; IAPH.

DR InterPro: IPR004825; Ins/IGF/relax.

DR PRINTS; PR00277; INSULINB.

DR SMART: SM00078; ILGF: 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism.  
 FT CHAIN 1 30 INSULIN B CHAIN.  
 FT CHAIN 1 30 INSULIN A CHAIN.  
 FT NON\_CONS 30 31  
 FT CHAIN 31 51 INSULIN A CHAIN.  
 FT CHAIN 31 51 INTERCHAIN.  
 FT DISULFID 7 37  
 FT DISULFID 19 50  
 FT DISULFID 36 41  
 SQ SEQUENCE 51 AA: 5745 MW: 90075096A0A7DDD CRC64;  
 Query Match 44.7%; Score 262.5; DB 1; Length 51;  
 Best Local Similarity 90.4%; Pred. No. 2.4e-21;  
 Matches 47; Conservative 2; Mismatches 2; Indels 1; Gaps 1;  
 QY 56 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEOCCTSCISLYOLENYCN 107  
 Db 1 FVNHLCGSHLVEALYLVCGERGFFYTPKRA-GIVEOCCASVCSLYOLENYCN 51  
 RESULT 15  
 INS\_CANFA STANDARD: PRT; 110 AA.  
 ID INS\_CANFA  
 AC P01321;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Insulin precursor.  
 GN INS.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=83109071; PubMed=6296142;  
 RA Kwok S.C.M., Chan S.J., Steiner D.F.;  
 RT "Cloning and nucleotide sequence analysis of the dog insulin gene.  
 RT Coded amino acid sequence of canine preproinsulin predicts an  
 RT additional C-peptide fragment.";  
 RL J. Biol. Chem. 258:2357-2363(1983).  
 RN [2]  
 RN SEQUENCE OF 25-54 AND 90-110  
 RX MEDLINE=66160119; PubMed=5949593;  
 RA Smith L.F.;  
 RT "Species variation in the amino acid sequence of insulin.";  
 RL Am. J. Med. 40:662-666(1966).  
 CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
 CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: V00179; CAA23475.1; -  
 DR PIR: A92413; IPDG.  
 DR HSSP: P01317; IAPH.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR SMART: SM00078; ILGF: 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism; Signal.  
 FT SIGNAL 1 24

FT CHAIN 25 54 INSULIN B CHAIN.  
 FT PROPEP 57 87  
 FT CHAIN 90 110 INSULIN A CHAIN.  
 FT DISULFID 31 96 INTERCHAIN.  
 FT DISULFID 43 109 INTERCHAIN.  
 FT DISULFID 95 100  
 SQ SEQUENCE 110 AA: 12190 MW: A574791864A4FB98 CRC64;  
 Query Match 44.6%; Score 262; DB 1; Length 110;  
 Best Local Similarity 59.3%; Pred. No. 5.9e-21;  
 Matches 51; Conservative 0; Mismatches 1; Indels 14; Gaps 1;  
 QY 56 FVNHLCGSHLVEALYLVCGERGFFYTPKTRGIVEOCCTSCISLYOLENYCN 107  
 Db 25 FVNHLCGSHLVEALYLVCGERGFFYTPKARREVEDLQVRDVELAGAFGGGLQPIALEG 84  
 QY 86 ----RGIVEOCCTSCISLYOLENYCN 107  
 Db 85 ALQKRGIVEOCCTSCISLYOLENYCN 110

Search completed: September 16, 2003, 12:38:51  
 Job time : 9.32685 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: September 16, 2003, 12:34:00 : Search time 25.8132 Seconds  
(without alignments)  
1669.670 Million cell updates/sec

Title: us-09-423-100-6

Perfect score: 587

Sequence: 1 MFPTPLSLRLFDNAWLAHRAH.....IVEQCCTICSLSYLENYCN 107

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL23:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mhc:  
8: sp\_organelle:  
9: sp\_phase:  
10: sp\_plant:  
11: sp\_rodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_rvirus:  
16: sp\_bacteriap:  
17: sp\_archaeup:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	267	45.5	110	6 Q8HXV2	Q8HXV2 pongo pygma
2	266	45.3	110	11 Q91XI3	Q91XI3 spermophilu
3	251	42.8	110	6 Q8XNW6	Q8XNW6 felis silve
4	249	42.4	217	6 Q8XNE0	Q8XNE0 ateles geof
5	228	38.8	245	4 Q14644	Q14644 homo sapien
6	219.5	37.4	106	13 Q91807	Q91807 rana pipien
7	213	36.3	212	6 Q07368	Q07368 macaca mula
8	213	36.3	217	6 Q07367	Q07367 macaca mula
9	201.5	34.3	110	13 Q98TA8	Q98TA8 pantodon bu
10	201	34.2	217	6 Q07369	Q07369 macaca mula
11	197	33.6	217	4 Q14407	Q14407 homo sapien
12	195.5	33.3	108	13 Q9DD55	Q9DD55 brachydanio
13	195.5	33.3	108	13 Q90ZN4	Q90ZN4 catla catla
14	195	33.2	111	13 Q98TB0	Q98TB0 chitala chi
15	195	33.2	217	6 Q8XND9	Q8XND9 ateles geof
16	193.5	33.0	110	13 Q90ZY1	Q90ZY1 hiodon alos

17	191.5	32.6	111	13 Q98TA7	Q98TA7 osteoglossu
18	189.5	32.3	108	13 Q98TB1	Q98TB1 catostomus
19	187.5	31.9	87	13 Q98TA9	Q98TA9 gnathonemus
20	185.5	31.6	91	13 Q98TB2	Q98TB2 ambloplites
21	171.5	29.2	217	6 Q8MI74	Q8MI74 callithrix
22	161	27.4	216	11 Q70615	Q70615 spalax leuc
23	159.5	27.2	52	6 Q9TV91	Q9TV91 equus cabal
24	159.5	27.2	216	6 QMI73	QMI73 delphinus d
25	159.5	27.2	216	6 QHYE5	QHYE5 alluropoda
26	156	26.6	216	11 Q9R2C3	Q9R2C3 mus musculus
27	154	26.2	216	11 Q9JRM4	Q9JRM4 cavia porce
28	152	25.9	178	6 Q95MJ5	Q95MJ5 tarsius ban
29	149.5	25.5	204	6 Q95205	Q95205 ovis aries
30	149	25.4	132	13 Q8AV14	Q8AV14 petromyzon
31	148.5	25.3	159	13 Q93607	Q93607 parolichthy
32	148.5	25.3	182	13 Q73720	Q73720 oreochromis
33	148.5	25.3	182	13 Q42289	Q42289 oreochromis
34	148.5	25.3	182	13 Q79824	Q79824 oreochromis
35	148.5	25.3	185	13 Q57436	Q57436 paralichthy
36	148.5	25.3	186	13 Q93527	Q93527 paralichthy
37	148	25.2	210	13 Q91443	Q91443 squalus aca
38	147.5	25.1	116	13 Q91161	Q91161 oncorhynch
39	147.5	25.1	117	13 Q91476	Q91476 salmo salar
40	147.5	25.1	145	13 Q91475	Q91475 salmo salar
41	147.5	25.1	149	13 Q91231	Q91231 oncorhynch
42	147.5	25.1	155	13 Q91162	Q91162 oncorhynch
43	147.5	25.1	161	13 Q91230	Q91230 oncorhynch
44	147.5	25.1	188	13 Q81268	Q81268 oncorhynch
45	147.5	25.1	188	13 Q91965	Q91965 oncorhynch

## ALIGNMENTS

RESULT 1

Q8HXV2 ID Q8HXV2 PRELIMINARY; PRT: 110 AA.  
AC Q8HXV2:  
DT 01-MAR-2003 (TremBLrel. 23, Created)  
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)  
DE Insulin precursor.  
GN INS.  
OS Pongo pygmaeus (orangutan)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catherhini; Eumetardae; Pongo.  
OX NCBI\_TaxID=9550;  
RN [1];  
RP SEQUENCE FROM N.A.  
RA Stead J.D.H., Jeffreys A.J.;  
RT "Haplotype diversity at the insulin region."  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY137503; AAN06937.1;  
SQ SEQUENCE 110 AA; 12038 MW; 22D2B32B94F520F8 CRC64;

Query Match 45.5%; Score 267; DB 6; Length 110;  
Best Loca. Similarity 60.5%; Pred. No. 6.6e-24;  
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

Qy	56	FVNHLCGSHLVPEALYLVCGERGFYTPKT-----	85
Db	25	FVNHLCGSHLVPEALYLVCGERGFYTPKT-----	84
Qy	86	----RGIVEQCCTICSLSYLENYCN 107	
Db	85	SLQKRGIVEQCCTICSLSYLENYCN 110	

RESULT 2

Q91XI3 ID Q91XI3 PRELIMINARY; PRT: 110 AA.  
AC Q91XI3:  
DT 01-DEC-2001 (TremBLrel. 19, Created)

```

DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Insulin.
OS Sperophilus tridecemlineatus (Thirteen-lined ground squirrel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Scluridae; Sclurinae;
OC Sperophilus.
OX NCBI_TaxID=43179;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RT Tredrea M.M., Buck M.J., Guhaniyogi J., Squire T.L., Andrews M.T.;
RA "Regulation of PK4 expression in a hibernating mammal.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL: AF038604; AA072558.1; -.
DR HSSP: P01308; ILNP.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00277; INSULINB.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: P10262; INSULIN; 1.
SQ SEQUENCE 110 AA; 12004 MW; 4511768D66228EE5 CRC64;

Query Match 45.3%; Score 266; DB 11; Length 110;
Best Local Similarity 57.4%; Pred. No. 8.7e-24;
Matches 54; Conservative 1; Mismatches 3; Indels 36; Gaps 2;

QY 50 LCTGP-RPVNHLCGSHLVALYVCGERGFFYTPKT----- 85
DB 17 LQDPAQAFVNHLCGSHLVALYVCGERGFFYTPKSRHREVEQGGVGGPGAGL 76

QY 86 -----RGIVEQCCTSICSLYQLENYCN 107
DB 77 PQLALEMALQKRGIVEQCCTSICSLYQLENYCN 110

RESULT 3
Q8WNW6
ID 08WNW6 PRELIMINARY; PRI: 110 AA.
AC 08WNW6;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Preproinsulin.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RT Okamoto S., Morimatsu M.;
RA "Cat insulin.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL: AB043535; BA084110.1; -.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00277; INSULINB.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
SQ SEQUENCE 110 AA; 12069 MW; 95F86E170C7BECA4 CRC64;

Query Match 42.8%; Score 251; DB 6; Length 110;
Best Local Similarity 55.8%; Pred. No. 5.3e-22;
Matches 48; Conservative 2; Mismatches 22; Indels 34; Gaps 1;

QY 56 FVNHLCGSHLVALYVCGERGFFYTPKT----- 85
DB 25 FVNHLCGSHLVALYVCGERGFFYTPKARPEADLQKDAELCEAPGAGGLQPSALEA 84

```

```

QY 86 -----RGIVEQCCTSICSLYQLENYCN 107
DB 85 PLOKRGIVECCASVCSLYQLEHYCN 110

RESULT 4
Q8WNEO
ID 08WNEO PRELIMINARY; PRI: 217 AA.
AC 08WNEO;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Growth hormone.
GN GH-N.
OS Ateles geoffroyi (Black-handed spider monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Ateleinae; Ateles.
OX NCBI_TaxID=9509;
RN [1]
RP SEQUENCE FROM N.A.
RA Revol A., Esquivel D., Santiago D., Barrera Saldana H.;
RT "Independent duplication of the growth hormone gene in three
RT Anthropoides lineages.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF374234; AAL72286.1; -.
DR InterPro: IPR001400; Somatotrophin.
DR Pfam: PF00103; hormone; 1.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
SQ SEQUENCE 217 AA; 24894 MW; 425829FF41EEAAE6 CRC64;

Query Match 42.4%; Score 249; DB 6; Length 217;
Best Local Similarity 97.9%; Pred. No. 2e-21;
Matches 47; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTIPISRFLFDNAMLRANKLHQLAFDTYQFFEEAYIPKEQKYSFLONP 49
DB 27 FPTIPISRLLDNAMLRANKLHQLAFDTYQFFEEAYIPKEQKYSFLONP 74

RESULT 5
Q14644
ID 014644 PRELIMINARY; PRI: 245 AA.
AC 014644;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Placental growth hormone isoform hGH V4 precursor.
GN HGH-V.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Term placenta;
RX MEDLINE=98373737; PubMed=9709963;
RA Boguszewski C.L., Svensson P.A., Jansson T., Clark R.,
RA Carlsson L.M.S., Carlsson B.;
RT "Cloning of two novel growth hormone transcripts expressed in human
RT placenta.";
RL J. Clin. Endocrinol. Metab. 81:2878-2885(1998).
DR EMBL: AF006061; AAB71829.1;
DR HSSP: P01241; 1A22.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
KW Signal.
FT SIGNAL 1 26 POTENTIAL.
SQ SEQUENCE 245 AA; 27101 MW; 14CC7F8CD75D91C8 CRC64;

Query Match 38.8%; Score 228; DB 4; Length 245;

```

DR EMBL: L16553; AAA18840.1; -

DR HSSP: P01241; IAXI.

DR InterPro: IPR001400; Somatotropin.

DR Pfam: PF00103; hormone; 1.

DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.

FT NON\_TER 1 1

SO SEQUENCE 212 AA; 24525 MW; 27BC91106256E6F5 CRC64;

Query Match 36.3%; Score 213; DB 6; Length 212;

Best Local Similarity 78.7%; Pred. No. 3, 8e-17;

Matches 37; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

QY 3 PTIPLSRLEFDNAMLRAHRLHQLAFDITYQFEFEAYIPKEOKYSFLGNP 49

DB 23 PSVPLSRLEFDHAMIQAHRLHQLAFDITYQFEFEAYIPKEKKHSLMNP 69

RESULT 8

Q07367 PRELIMINARY: PKT: 217 AA.

ID Q07367

AC Q07367

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE Choriionic somatomamotropin-1.

OS Macaca mulatta (Rhesus macaque).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;

OC Cercopitheciinae; Macaca.

OX NCBI\_TaxID=9544;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Midpregnancy placenta;

RX MEDLINE=94008724; PubMed=8404617;

RA Golos T.G., Durning M., Fisher J.M., Fowler P.D.;

RT "Cloning of four growth hormone/chorionic somatomamotropin-related

RT complementary deoxyribonucleic acids differentially expressed during

RT pregnancy in the rhesus m key placenta.";

RL Endocrinology 133:1744-17 1993).

DR EMBL: L16552; AAA18839.1;

DR HSSP: P01241; IAXI.

DR InterPro: IPR001400; Somatotropin.

DR Pfam: PF00103; hormone; 1.

DR PRINTS: PR00836; SOMATOTROPIN.

DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.

SO SEQUENCE 217 AA; 24942 MW; FF5AA8915131F2BC CRC64;

Query Match 46.3%; Score 213; DB 6; Length 217;

Best Local Similarity 78.7%; Pred. No. 3, 8e-17;

Matches 37; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

QY 3 PTIPLSRLEFDNAMLRAHRLHQLAFDITYQFEFEAYIPKEOKYSFLGNP 49

DB 28 PSVPLSRLEFDHAMIQAHRLHQLAFDITYQFEFEAYIPKEKKHSLMNP 74

RESULT 9

Q98TA8 PRELIMINARY: PKT: 110 AA.

ID Q98TA8

AC Q98TA8;

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE Preproinsulin.

OS Pantodon buchholzi (Butterflyfish).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;

OC Osteoglossiformes; Pantodontidae; Pantodon.

OX NCBI\_TaxID=8276;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21203577; PubMed=11306171;

RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;

RT \*Molecular cloning of preproinsulin cDNAs from several  
 RL osteoclasts and a cyprinid.  
 CC Mol. Cell. Endocrinol. 174:51-58(2001).  
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

DR EMBL: AF199588; AAK28712.1; -;  
 DR HSSP: P01308; I1H5.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 DR SEQUENCE 110 AA; 12324 MW; BDECCD659D872E06 CRC64;

Query Match 34.3%; Score 201.5; DB 13; Length 110;  
 Best Local Similarity 43.5%; Pred. No. 4e-16;  
 Matches 37; Conservative 8; Mismatches 5; Indels 35; Gaps 1;

QY 58 NQHLGGSHLVYLCGRCFFVTPKT----- 85

DB 26 SOHLGGSHLVYLCGRCFFVTPKT----- 85

QY 86 ---RGIVEQCCTSCSLYQLENYCN 107

DB 86 KVKRGIVEQCCHPCNIFDLQNYCN 110

## RESULT 10

ID Q07369 PRELIMINARY; PRT: 217 AA.

AC Q07369;

DI 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Chorionic somatomotropin-3.

OS Macaca mulatta (Rhesus macaque).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;

OC Cercopithecoidea; Macaca.

OX NCBI\_TaxID=9544;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Midpregnancy placenta;

RX MEDLINE=94008724; PubMed=8404617;

RA Golos T.G., Durning M., Fisher J.M., Fowler P.D.;

RT Cloning of four growth hormone/chorionic somatomotropin-related

RT complementary deoxyribonucleic acids differentially expressed during

RT pregnancy in the rhesus monkey placenta.;

RL Endocrinology 133:1744-1752(1993).

DR EMBL: L16554; AAA18841.1; -;

DR HSSP: P01241; IAXI.

DR InterPro: IPR001400; Somatotropin.

DR Pfam: PF00103; hormone; 1.

DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.

DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.

DR SEQUENCE 217 AA; 24874 MW; FIER6AFDBA1B185 CRC64;

Query Match 34.2%; Score 201; DB 6; Length 217;  
 Best Local Similarity 74.5%; Pred. No. 1e-15;  
 Matches 35; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 3 PIPSLRFDNMLRAHRLHQLAFDTYQEFEEAIPKEQKYSFLONP 49

DB 28 PSVPLSLRFDNMLRAHRLHQLAFDTYQEFEEAIPKEKXSLMGNP 74

## RESULT 11

ID Q14407 PRELIMINARY; PRT: 217 AA.

AC Q14407;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Chorionic somatomotropin CS-2 (Chorionic somatomotropin hormone

DE Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89307277; PubMed=2744760;  
 RA Chen E.Y., Liao Y.C., Smith D.H., Barrera-Saldana H.A., Gellinas R.E.,  
 RA Seeburg P.H.;  
 RT \*The human growth hormone locus: nucleotide sequence, biology, and  
 RT evolution.;

RL Genomics 4:479-497(1989).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=91102558; PubMed=1158;

RA Vnencak-Jones C.L., Phillips J.A. III;

RT \*Hot spots for growth hormone gene deletions in homologous regions

RT outside of Alu repeats.;

RL Science 250:1745-1748(1990).

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE=Placenta;

RA Strausberg R.;

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RC TISSUE=Placenta;

RA Strausberg R.;

RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: J01071; AA52553.1; -;

DR EMBL: BC022044; AAH22044.1; -;

DR EMBL: BC035965; AAH35965.1; -;

DR HSSP: P01241; IAX2.

DR InterPro: IPR001400; Somatotropin.

DR Pfam: PF00103; hormone; 1.

DR PROSITE: PS00836; SOMATOTROPIN

DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.

DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.

DR SEQUENCE 217 AA; 24994 MW; 39FAACDDB6B62E951 CRC64;

Query Match 33.6%; Score 197; DB 4; Length 217;  
 Best Local Similarity 80.0%; Pred. No. 3e-15;  
 Matches 36; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 4 TIPSRLFDNMLRAHRLHQLAFDTYQEFEEAIPKEQKYSFLON 48

DB 29 TVPLSLRFDNMLRAHRLHQLAFDTYQEFEEAIPKEQKYSFLHD 73

## RESULT 12

Q9DDE5 PRELIMINARY; PRT: 108 AA.

ID Q9DDE5;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Insulin precursor.

GN INS.

OS Brachydanio rerio (Zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Danio.

OX NCBI\_TaxID=7955;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99425190; PubMed=10495291;

RA Argenton F., Zecchin E., Bortolussi M.;

RT \*Early appearance of pancreatic hormone-expressing cells in the

RT zebrafish embryo.;

RL Mech. Dev. 87:217-221(1999).

CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

```

DR EMBL: AJ237750; CAC20109.1;
DR HSSP: P01308; ILPH;
DR ZFIN: ZDB-GENE-980526-110; Ins.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00277; INSULINB.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
KW Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 53 INSULIN B CHAIN.
FT CHAIN 86 108 INSULIN A CHAIN.
SQ SEQUENCE 108 AA; 11904 MW; 3195289E72AD6D25 CRC64;

Query Match 33.3%; Score 195.5; DB 13; Length 108;
Best Local Similarity 45.1%; Pred. No. 2e-15;
Matches 37; Conservative 5; Mismatches 7; Indels 33; Gaps 1;

QY 59 QHLGSHLYEALVLCGERGFYTPK-----T 85
D 27 QHLGSHLYDALVLCGGPTGFYFNPKROVEPLLGFLPKSAQETEVADFADFKDHAELIRK 86

QY 86 RGIVEOCCTSICSLYLENYCN 107
D 87 RGIVEOCCCHKPCIFELQNYCN 108

RESULT 13
Q902N4 PRELIMINARY; PRT: 108 AA.
AC Q902N4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Preproinsulin.
OS Catla catla (Catla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Catla.
OX NCBI_TaxID=72446;
RN [1]
RP SEQUENCE FROM N.A.
RA Bhattacharya S., Roy S.S., Dasgupta S., Ravikumar L., Mukherjee M.,
RA Bandyopadhyaya I., Wakabayashi K.;
RT "A new cell secreting insulin.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL: AF373021; AAK51558.1;
DR HSSP: P01308; ILNP.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00277; INSULINB.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
SQ SEQUENCE 108 AA; 11881 MW; D713026E22EF5D59 CRC64;

Query Match 33.3%; Score 195.5; DB 13; Length 108;
Best Local Similarity 45.1%; Pred. No. 2e-15;
Matches 37; Conservative 5; Mismatches 7; Indels 33; Gaps 1;

QY 59 QHLGSHLYEALVLCGERGFYTPK-----T 85
D 27 QHLGSHLYDALVLCGGPTGFYFNPKROVDPLMGFLPKSAQETEVADFADFKDHAELIRK 86

QY 86 RGIVEOCCTSICSLYLENYCN 107
D 87 RGIVEOCCCHKPCIFELQNYCN 108

RESULT 14
Q98TB0 PRELIMINARY; PRT: 111 AA.
ID Q98TB0
QY 2 FTPIPLSRFDNAHLRAHLHQLAFDTYQFFEFAYIPKQKYSFLONP 49

```

```

AC Q98TB0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Preproinsulin (Fragment).
OS Chitala chitala (Clown knifefish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
OC Osteoglossiformes; Notopteridae; Chitala.
OX NCBI_TaxID=112153;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE: 21203577; PubMed: 110171;
RA Al-Mahrouki A.A., Irwin D.M., Ibrahim L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
RT osteoglossomorphs and a cyprinid.";
RL Mol. Cell. Endocrinol. 174:58 (2001).
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL: AF199586; AAK28710.1;
DR HSSP: P01308; ILPH.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR SMART: SM00078; IIGF; 1.
DR NON_TER 111
FT NON_TER 111
SQ SEQUENCE 111 AA; 12483 MW; 247CA4431376329F CRC64;

Query Match 33.2%; Score 195; DB 13; Length 111;
Best Local Similarity 44.2%; Pred. No. 2.4e-15;
Matches 38; Conservative 3; Mismatches 9; Indels 46; Gaps 1;

QY 58 NQHLGSHLYEALVLCGERGFYTPK-----84
D 26 NQHLGSHLYEALVLCGERGFYFNPKMDKROAEPLLGFLSPKSGLENEVDYEPFKDQKD 85

QY 85 ---TRGIVEOCCTSICSLYLENYCN 107
D 86 VKMKRGIVEOCCCHRPCNIFDQNYCN 111

RESULT 15
Q8WND9 PRELIMINARY; PRT: 217 AA.
ID Q8WND9
AC Q8WND9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Growth hormone.
OS GH-V.
OC Atles Geoffroyi (Black-handed spider monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cetiidae; Ateles.
OX NCBI_TaxID=9509;
RN [1]
RP SEQUENCE FROM N.A.
RA Revol A., Esquivel D., Santiago D., Barrera-Saldana H.
RT "Independent duplication of the growth hormone gene in three
RT Anthropoidean lineages.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF374235; AAL72287.1;
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
SQ SEQUENCE 217 AA; 25293 MW; 741745A1B75C053E CRC64;

Query Match 33.2%; Score 195; DB 6; Length 217;
Best Local Similarity 77.1%; Pred. No. 5.2e-15;
Matches 37; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 2 FTPIPLSRFDNAHLRAHLHQLAFDTYQFFEFAYIPKQKYSFLONP 49

```

Db 27 FPRIPLSRFGDAMLRAHQLHQAFTYQLELENCIPKKQKYEFLRNP 74

Search completed: September 16, 2003, 12:40:00  
Job time : 27.8132 secs



PT particularly for the production of human insulin  
 XX  
 PS Claim 14; Page 30-31; 46pp; English.

CC This sequence represents a chimeric protein, which contains an  
 CC N-terminal fragment of human growth hormone (hGH) of the sequence given  
 CC in AAY42856, a cleavable peptide linker (AAY42857), and a human insulin  
 CC precursor comprising insulin A and B chains (AAY42859). The hGH portion  
 CC of the chimeric protein acts as an intramolecular chaperone (IMC) for  
 CC the insulin precursor, enabling it to fold correctly. The cleavable  
 CC peptide linker has a C-terminal Arg residue which enables the hGH portion  
 CC of the chimeric protein to be removed after folding has taken place.  
 CC Production of recombinant human insulin via an hGH-proinsulin chimeric  
 CC protein can provide human insulin with correctly linked cysteine bridges  
 CC with fewer necessary procedural steps, and hence resulting in a higher  
 CC yield of human insulin. The IMC sequences not only protect insulin  
 CC sequences from intracellular degradation by a microorganism host, but  
 CC also promote the folding of the fused insulin precursor, facilitate the  
 CC solubility of the fusion protein and decrease the intermolecular  
 CC interactions among the fusion proteins, thus allowing folding of the  
 CC fused insulin precursor at commercially useful high concentrations. The  
 CC procedural steps of cyanogen bromide cleavage, oxidative sulphytolysis  
 CC and related purification steps can thus be eliminated, along with the use  
 CC of high concentrations of mercaptan or the use of hydrophobic absorbent  
 CC resins.

XX Sequence 150 AA;  
 SQ Query Match 100.0%; Score 797; DB 20; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 1e-42;  
 Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFPTPLSRFLDNAMLRHLRLHQLAFDTYQEFEEAYIPKEQKYSFLONPQTSLSFSSESIP 60  
 DB 1 MFPTPLSRFLDNAMLRHLRLHQLAFDTYQEFEEAYIPKEQKYSFLONPQTSLSFSSESIP 60  
 QY 61 TPSNREETQOKSNLELRISLLLIQSWLEPVQLGTGPRFVNOHLCGSHLYEALYLVCGER 120  
 DB 61 TPSNREETQOKSNLELRISLLLIQSWLEPVQLGTGPRFVNOHLCGSHLYEALYLVCGER 120  
 QY 121 GFYTPKTRGIVEQCCTSCISLYOLENYCN 150  
 DB 121 GFYTPKTRGIVEQCCTSCISLYOLENYCN 150

RESULT 2  
 AAY42860  
 ID AAY42860 standard; protein: 107 AA.

XX AAY42860;  
 AC  
 DT 19-JAN-2000 (first entry)

XX hGH-mini-proinsulin chimeric protein.

DE Insulin; precursor; growth hormone; chaperone; intramolecular;  
 KW folding; conformation; chimeric protein; cleavable; recombinant;  
 KW production; yield.

XX Synthetic.  
 OS Homo sapiens.

XX W09950302-A1.

XX 07-OCT-1999.

XX 31-MAR-1998; 98WO-CN00052.

XX 31-MAR-1998; 98WO-CN00052.

XX (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.

XX Gan 2;

XX WPI; 1999-610839/52.  
 DR  
 XX  
 PT New chimeric proteins containing human growth hormone fragment, used  
 particularly for the production of human insulin  
 XX  
 PS Claim 13; Page 30; 46pp; English.

XX This sequence represents a chimeric protein, hGH-mini-proinsulin.  
 CC This chimeric protein contains an N-terminal fragment of human growth  
 CC hormone (hGH) of the sequence given in AAY42855, a cleavable peptide  
 CC linker (AAY42857), and a human insulin precursor comprising insulin  
 CC A and B chains (AAY42859). The hGH portion of the chimeric protein acts  
 CC as an intramolecular chaperone (IMC) for the insulin precursor,  
 CC enabling it to fold correctly. The cleavable peptide linker has a  
 CC C-terminal Arg residue which enables the hGH portion of the  
 CC chimeric protein to be removed after folding has taken place. Production  
 CC of recombinant human insulin via an hGH-proinsulin chimeric protein can  
 CC provide human insulin with correctly linked cysteine bridges with  
 CC fewer necessary procedural steps, and hence resulting in a higher yield  
 CC of human insulin. The IMC sequences not only protect insulin sequences  
 CC from intracellular degradation by a microorganism host, but also promote  
 CC the folding of the fused insulin precursor, facilitate the solubility of  
 CC the fusion protein and decrease the intermolecular interactions among  
 CC the fusion proteins, thus allowing folding of the fused insulin precursor  
 CC at commercially useful high concentrations. The procedural steps of  
 CC cyanogen bromide cleavage, oxidative sulphytolysis and related  
 CC purification steps can thus be eliminated, along with the use of high  
 CC concentrations of mercaptan or the use of hydrophobic absorbent resins.

XX Sequence 107 AA;

Query Match 69.7%; Score 555.5; DB 20; Length 107;  
 Best Local Similarity 71.3%; Pred. No. 8e-28;  
 Matches 107; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MFPTPLSRFLDNAMLRHLRLHQLAFDTYQEFEEAYIPKEQKYSFLONPQTSLSFSSESIP 60  
 DB 1 MFPTPLSRFLDNAMLRHLRLHQLAFDTYQEFEEAYIPKEQKYSFLONPQTSLSFSSESIP 60  
 QY 61 TPSNREETQOKSNLELRISLLLIQSWLEPVQLGTGPRFVNOHLCGSHLYEALYLVCGER 120  
 DB 50  
 QY 121 GFYTPKTRGIVEQCCTSCISLYOLENYCN 150  
 DB 78 GFYTPKTRGIVEQCCTSCISLYOLENYCN 150

RESULT 3

AAY42856

ID AAY42856 standard; protein: 92 AA.

XX AAY42856;

XX 19-JAN-2000 (first entry)

DE Human growth hormone (hGH) N-terminal fragment #2.

KW Growth hormone; chaperone; intramolecular; insulin; precursor;  
 KW folding; conformation; chimeric protein; cleavable; recombinant;  
 KW production; yield.

XX Homo sapiens.

XX W09950302-A1.

XX 07-OCT-1999.

XX 31-MAR-1998; 98WO-CN00052.

XX 31-MAR-1998; 98WO-CN00052.



PA (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.  
 XX  
 PI Gan 2;  
 XX  
 XX WPI; 1999-610839/52.  
 XX  
 PT New chimeric proteins containing human growth hormone fragment, used  
 PT particularly for the production of human insulin  
 XX  
 XX Claim 5; Page 28; 46pp; English.  
 XX  
 CC This sequence represents an N-terminal fragment of human growth hormone  
 CC (hGH) which is a component of a chimeric protein (AA42861) which also  
 CC contains a human insulin precursor (AA42859). The hGH portion of the  
 CC chimeric protein acts as an intramolecular chaperone (IMC) for the  
 CC insulin precursor, enabling it to fold correctly. A cleavable peptide  
 CC linker with a C-terminal Arg residue (AA42857) enables the hGH portion  
 CC of the chimeric protein to be removed after folding has taken place.  
 CC Production of recombinant human insulin via an hGH-proinsulin chimeric  
 CC protein can provide human insulin with correctly linked cysteine bridges  
 CC with fewer necessary procedural steps, and hence resulting in a higher  
 CC yield of human insulin. The IMC sequences not only protect insulin  
 CC sequences from intracellular degradation by a microorganism host, but  
 CC also promote the folding of the fused insulin precursor, facilitate the  
 CC solubility of the fusion protein and decrease the intermolecular  
 CC interactions among the fusion proteins, thus allowing folding of the  
 CC fused insulin precursor at commercially useful high concentrations. The  
 CC procedural steps of cyanogen bromide cleavage, oxidative sulphytolysis  
 CC and related purification steps can thus be eliminated, along with the  
 CC use of high concentrations of mercaptan or the use of hydrophobic  
 CC absorbent resins.  
 XX  
 XX Sequence 92 AA;  
 SQ  
 Query Match 59.0%; Score 470; DB 20; Length 92;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-22;  
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MFPTPLSLRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPTSLSFSESIP 60  
 DB 1 MFPTPLSLRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPTSLSFSESIP 60  
 QY 61 TPSNRETOOKSNLELLRISLLLIQSWLEPVQ 92  
 DB 61 TPSNRETOOKSNLELLRISLLLIQSWLEPVQ 92  
 RESULT 4  
 AAW92265  
 XX AAW92265 standard; Protein: 134 AA.  
 XX  
 AC AAW92265;  
 XX  
 XX 08-JUN-1999 (first entry)  
 XX  
 DE Human anti-angiogenic peptide 16K hGH Met-1prol33.  
 XX  
 KW Human; anti-angiogenic; prolactin; placental lactogen; hPL; angiogenesis;  
 KW growth hormone; hGH; hGH-V; capillary endothelial cell proliferation;  
 KW placental vascularisation; pregnancy; treatment; angiogenic disease;  
 KW tumour; inhibitor; malignant; angiofibroma; arteriovenous malformation;  
 KW arthritis; atherosclerotic plaques; corneal graft neovascularisation;  
 KW wound healing; proliferative retinopathy; macular degeneration; trachoma;  
 KW psoriasis; fibroplasia; scleroderma; Kaposi's sarcoma; vascular adhesion;  
 KW ulcer; leukaemia; reproductive disorder; contraceptive agent;  
 KW gene therapy; pre-eclampsia; intrauterine growth retardation;  
 KW placental dysfunction.  
 XX  
 OS Homo sapiens.  
 XX  
 XX W09851323-A1.  
 XX

PD 19-NOV-1998.  
 XX  
 PF 12-MAY-1998; 98WO-US09691.  
 XX  
 PR 13-MAY-1997; 97US-0046394.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Martial JA, Struman I, Taylor R, Weiner RI;  
 XX  
 DR WPI; 1999-045192/04.  
 DR N-PSDB; AAX01707.  
 XX  
 XX New anti-angiogenic peptides - comprise N-terminal fragments of  
 PT human placental lactogen, human growth hormone, growth hormone  
 PT variant or human prolactin  
 XX  
 PS Claim 4; Page 49-50; 87pp; English.  
 XX  
 CC This invention describes novel human anti-angiogenic peptides derived  
 CC from 10 to 150 consecutive amino acids selected from the N-terminal end  
 CC of human placental lactogen, hPL, human growth hormone (hGH), growth  
 CC hormone variant (hGH-V), or human prolactin. Such peptides (i) inhibit  
 CC capillary endothelial cell proliferation and organisation (ii) inhibit  
 CC angiogenesis in chick chorioallantoic membrane and (iii) binds to at  
 CC least one specific receptor which does not bind an intact full length  
 CC hGH, hPL, prolactin or hGH-V. The invention also describes a method for  
 CC diagnosing a probable abnormality of placental vascularisation during  
 CC pregnancy. The peptides can be used for treating an angiogenic disease in  
 CC a subject, for inhibiting tumour formation or growth in a patient or for  
 CC modulating vascularisation of a patient's placenta. In particular, the  
 CC peptides can be used for preventing or treating e.g. malignant tumours,  
 CC angiofibroma, arteriovenous malformation, arthritic such as rheumatoid  
 CC arthritis, atherosclerotic plaques, corneal graft neovascularisation,  
 CC delayed wound healing, proliferative retinopathy such as diabetic  
 CC retinopathy, macular degeneration, granulations such as those occurring  
 CC in haemophilic joints, inappropriate vascularisation in wound healing  
 CC such as hypertrophic scars or keloid scars, neovascular glaucoma, ocular  
 CC tumour, uveitis, non-union fractures, Osler-Weber syndrome, psoriasis,  
 CC pyogenic glaucoma, retrolental fibroplasia, scleroderma, solid tumours,  
 CC Kaposi's sarcoma, trachoma, vascular adhesions, chronic varicose ulcers,  
 CC leukaemia, and reproductive disorders such as follicular and luteal cysts  
 CC and choriocarcinoma. They can also be used as contraceptive agents. DNA  
 CC encoding the peptides can be used in gene therapy. The measurement of  
 CC abnormal levels of N-terminal fragments of hGH, hGH-V, prolactin or hPL  
 CC can be used in assays for impairment of vascular development associated  
 CC with pre-eclampsia, intrauterine growth retardation, and placental  
 CC dysfunction.  
 XX  
 XX Sequence 134 AA;  
 SQ  
 Query Match 59.0%; Score 470; DB 20; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 2e-22;  
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MFPTPLSLRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPTSLSFSESIP 60  
 DB 1 MFPTPLSLRFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPTSLSFSESIP 60  
 QY 61 TPSNRETOOKSNLELLRISLLLIQSWLEPVQ 92  
 DB 61 TPSNRETOOKSNLELLRISLLLIQSWLEPVQ 92  
 RESULT 5  
 ABG94861  
 ID ABG94861 standard; Protein: 191 AA.  
 XX  
 AC ABG94861;  
 XX  
 XX 03-DEC-2002 (first entry)  
 XX  
 DE Human growth hormone mutant hPRL (111-129).

XX Growth hormone; placental lactogen; prolactin; active domain; hGH;  
 KW structure-function relationship; segment-substituted polypeptide;  
 KW mutant; mutin.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN US6428954-B1.  
 XX  
 PD 06-AUG-2002.  
 XX  
 PF 06-JUN-1995; 95US-0483039.  
 XX  
 PR 26-OCT-1989; 89US-0428066.  
 PR 27-APR-1992; 92US-0875204.  
 PR 13-OCT-1992; 92US-0960227.  
 PR 02-FEB-1994; 94US-0190723.  
 PR 28-OCT-1988; 88US-0264611.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Wells JA, Cunningham BC;  
 XX  
 DR WPI: 2002-696875/75.  
 XX  
 PT Identifying active domains within cloned polypeptides of known amino  
 PT acid sequence by substituting analog segments into the parent  
 PT polypeptide is useful to determine the relationship between structure  
 PT and function.  
 XX  
 PS Example 1; Page -: 86pp; English.  
 XX  
 CC The invention relates to identifying an unknown active domain in a region  
 CC of known amino acid sequence in a parent polypeptide e.g. human growth  
 CC hormone (hGH) which has been cloned and has a pre-identified biological  
 CC activity, where the active domain interacts with a target when the parent  
 CC polypeptide is in its native-folded form and the interaction is  
 CC responsible for the biological activity comprising: (a) comparing the  
 CC amino acid sequence or polypeptide structure in the region of known amino  
 CC acid sequence of hGH with the amino acid sequence or polypeptide  
 CC structure in a region of known amino acid sequence of an analogue  
 CC polypeptide (e.g. prolactin, placental lactogen or porcine growth  
 CC hormone) which has at least 15% homology with hGH alpha-carbon  
 CC coordinates within about 2-3.5 angstroms of hGH alpha-carbon  
 CC for about 60% of the analogue sequence, where any interaction of the  
 CC analogue with the target is different from target interaction with hGH;  
 CC (b) substituting DNA encoding an analogous polypeptide segment from the  
 CC analogue into DNA encoding the full length hGH, and expressing a  
 CC segment-substituted polypeptide; (c) contacting the segment-substituted  
 CC polypeptide with the target to determine interaction; (d) repeating steps  
 CC (b) and (c) with a second analogous polypeptide segment; and  
 CC (e) comparing the difference between activity of the first and second  
 CC segment-substituted polypeptides as an indication of the location of  
 CC the unknown active domain in hGH. The method is useful for determining  
 CC the relationship between structure and function of known polypeptide  
 CC sequences. The present sequence is that of human growth hormone  
 CC mutant substituted with residues from an hGH analogue (prolactin,  
 CC placental lactogen or porcine growth hormone).  
 CC Note: The present sequence is not shown in the specification but was  
 CC created by the indexer using the mature hGH sequence and information  
 CC contained in the specification.  
 XX  
 SQ Sequence 191 AA;

Query Match 59.0%; Score 470; DB 23; Length 191;  
 Best Local Similarity 69.2%; Pred. No. 2.6e-22;  
 Matches 101; Conservative 8; Mismatches 19; Indels 18; Gaps 3;

QY 2 FPTPLSLFDNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLSFSSEIPT 61  
 |||||  
 DB 1 FPTPLSLFDNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLSFSSEIPT 60

QY 62 PSNREETOQKSNLELLRLISLLLIQSWLEPVLGTGPRFVNOHLCGS-----H 108  
 |||||  
 DB 61 PSNREETOQKSNLELLRLISLLLIQSWLEPVLGTGPRFVNOHLCGS-----H 119  
 |||||  
 QY 109 LVEALYLVGCGERGFYTPKTRGIVEQ 134  
 |||||  
 DB 120 LIEGLMLILSDG-----SPRT; IFKQ 141  
 |||||  
 RESULT 6  
 AAP90129 standard: protein; 192 AA.  
 ID AAP90129;  
 AC AAP90129;  
 XX 25-MAR-2003 (updated)  
 DT 06-FEB-1996 (revised)  
 DT 01-NOV-1989 (first entry)  
 XX Human growth hormone.  
 DE Human growth hormone; fusion protein; recombinant  
 KW vector.  
 KW Homo sapiens (Human).  
 OS JP01144981-A.  
 PN 07-JUN-1989.  
 PD 02-DEC-1987; 87JP-0304937.  
 PF 02-DEC-1987; 87JP-0304937.  
 XX (WAKI ) WAKUNAGA SEIYAKU KK.  
 PA WPI: 1989-209284/29.  
 DR N-PSDB; AAN90269.  
 XX Recombinant vector contg. fus. protein - consisting of human  
 PT growth hormone or deriv. lig. 1 to foreign protein, for stability  
 PT and high yield.  
 XX Disclosure: Fig 1; 19pp; Japanese.  
 CC The invention consists of a vector contg. a fusion protein which is  
 CC formed by ligating, downstream of a promoter, hGH or a deriv. (pref.  
 CC by substn. of Met-14 with leu) and a foreign protein.  
 CC Stability of the vector in the host is greatly increased so the  
 CC protein yield is higher.  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 XX Sequence 192 AA;  
 SQ  
 Query Match 58.5%; Score 466; DB 10; Length 192;  
 Best Local Similarity 70.5%; Pred. No. 4.6e-22;  
 Matches 103; Conservative 7; Mismatches 20; Indels 16; Gaps 4;  
 QY 1 MFPTPLSLFDNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLSFSSEIP 60  
 |||||  
 DB 1 MFPTPLSLFDNMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLSFSSEIP 60  
 |||||  
 QY 61 TPSNREETOQKSNLELLRLISLLLIQSWLEPVLGTGPRFVNOHLCGS-----H 110  
 |||||  
 DB 61 TPSNREETOQKSNLELLRLISLLLIQSWLEPVLGTGPRFVNOHLCGS-----H 119  
 |||||  
 QY 111 EALYLVGCGERGFYTPKTRGIVEQ 134  
 |||||  
 DB 120 EGIOTLMGRLEDC---SPRTGQIFKQ 142  
 |||||  
 RESULT 7  
 AAN92264

ID AAW92264 standard; Protein: 192 AA.

XX AAW92264;

XX 08-JUN-1999 (first entry)

XX Human anti-angiogenic peptide hGH Met-1p191.

XX Human; anti-angiogenic; prolactin; placental lactogen; hPL; angiogenesis;  
 KW growth hormone; hGH; hGH-V; capillary endothelial cell proliferation;  
 KW placental vasculature; pregnancy; treatment; angiogenic disease;  
 KW tumour; inhibitor; malignant; angiofibroma; arteriovenous malformation;  
 KW arthritis; atherosclerotic plaques; corneal graft neovascularisation;  
 KW wound healing; proliferative retinopathy; macular degeneration; trachoma;  
 KW granulation; glaucoma; ocular; uveitis; fracture; Osler-Weber syndrome;  
 KW psoriasis; fibroplasia; scleroderma; Kaposi's sarcoma; vascular adhesion;  
 KW ulcer; leukaemia; reproductive disorder; contraceptive agent;  
 KW gene therapy; pre-eclampsia; intrauterine growth retardation;  
 KW placental dysfunction.

XX Homo sapiens.

XX WO9851323-A1.

XX 19-NOV-1998.

XX 12-MAY-1998; 98WO-US09691.

XX 13-MAY-1997; 97US-0046394.

XX (REGC) UNIV CALIFORNIA.

XX Martial JA, Struman I, Taylor R, Weiner RI;

XX WPI: 1999-045192/04.

XX N-PSDB; AAX01706.

XX New anti-angiogenic peptides - comprise N-terminal fragments of  
 PT human placental lactogen, human growth hormone, growth hormone  
 PT variant or human prolactin

XX Example 3; Page 49; 87pp; English.

XX This invention describes novel human anti-angiogenic peptides derived  
 CC from 10 to 150 consecutive amino acids selected from the N-terminal end  
 CC of human placental lactogen (hPL), human growth hormone (hGH), growth  
 CC hormone variant (hGH-V), or human prolactin. Such peptides (i) inhibit  
 CC capillary endothelial cell proliferation and organisation (ii) inhibit  
 CC angiogenesis in chick chorioallantoic membrane and (iii) binds to at  
 CC least one specific receptor which does not bind an intact full length  
 CC hGH, hPL, prolactin or hGH-V. The invention also describes a method for  
 CC diagnosing a probable abnormality of placental vascularisation during  
 CC pregnancy. The peptides can be used for treating an angiogenic disease in  
 CC a subject, for inhibiting tumour formation or growth in a patient or for  
 CC modulating vascularisation of a patient's placenta. In particular, the  
 CC peptides can be used for preventing or treating e.g. malignant tumours,  
 CC angiofibroma, arteriovenous malformation, arthritic such as rheumatoid  
 CC arthritis, atherosclerotic plaques, corneal graft neovascularisation,  
 CC delayed wound healing, proliferative retinopathy such as diabetic  
 CC retinopathy, macular degeneration, granulations such as those occurring  
 CC in haemophilic joints, inappropriate vascularisation in wound healing  
 CC such as hypertrophic scars or keloid scars, neovascular glaucoma, ocular  
 CC tumour, uveitis, non-union fractures, Osler-Weber syndrome, psoriasis,  
 CC pyogenic glaucoma, retrolental fibroplasia, scleroderma, solid tumours,  
 CC Kaposi's sarcoma, trachoma, vascular adhesions, chronic varicose ulcers,  
 CC leukaemia, and reproductive disorders such as follicular and luteal cysts  
 CC and choriocarcinoma. They can also be used as contraceptive agents. DNA  
 CC encoding the peptides can be used in gene therapy. The measurement of  
 CC abnormal levels of N-terminal fragments of hGH, hGH-V, prolactin or hPL  
 CC can be used in assays for impairment of vascular development associated  
 CC with pre-eclampsia, intrauterine growth retardation, and placental  
 CC dysfunction.

XX

SQ Sequence 192 AA;

Query Match 58.5%; Score 466; DB 20; Length 192;

Best Local Similarity 70.5%; Pred. No. 4.6e-22;

Matches 103; Conservative 7; Mismatches 20; Indels 16; Gaps 4;

Oy 1 MFPTIPLSRLFDNAMLRAH; HQLAFTDYQEFEEAYIPKEQKYSFLONPOTLSFSSESIP 60

Db 1 MFPTIPLSRLFDNAMLRAH; HQLAFTDYQEFEEAYIPKEQKYSFLONPOTLSFSSESIP 60

Oy 61 TPSNRETOQKSNLELLRISLLLIQSWLEPVQIGTGRFFVNOHLGGS-----HLV 110

Db 61 TPSNRETOQKSNLELLRISLLLIQSWLEPVQIGTGRFFVNOHLGGS-----HLV 110

Oy 111 EALYLVGSG--ERGFYTPKTRGIVEQ 134

Db 120 EGITOTLMGRLEDG---SPRTQGIKQ 142

RESULT 8

AAP91041

ID AAP91041 standard; protein: 40 AA.

XX AAP91041;

XX 14-DEC-1989 (first entry).

XX Human growth hormone segment.

XX Human growth hormone; fusion protein; thrombin;

KW geriatric dementia; nervous disorders; human nerve factor.

XX Homo sapiens (human).

OS EP329175-A.

PN 23-AUG-1989.

XX 17-FEB-1989; 89EP-0102795.

XX 19-FEB-1988; 88JP-0035042.

XX (TOYJ) TOSOH CORP.

XX Ohtsuka E.

XX WPI: 1989-243092/34.

XX New human nerve growth factor gene encoding fusion protein

XX - having cleavage site for thrombin, useful for treating geriatric

XX dementia, etc.

XX Disclosure: page 21; 38pp; English.

XX Human growth hormone segment, used at the N-terminal of a fusion

XX protein, which contains a thrombin recognition site, and human beta nerve

XX growth factor (beta-NGF) at the C-terminal. Beta-NGF can be used to

XX control geriatric dementia and other nervous disorders, and can be

XX released from the fusion protein by incubation with thrombin (see

XX AAN90577-8, AAP91034, AAP91299).

XX SQ Sequence 140 AA;

Query Match 58.3%; Score 465; DB 10; Length 140;

Best Local Similarity 98.9%; Pred. No. 4.2e-22;

Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 MFPTIPLSRLFDNAMLRAH; HQLAFTDYQEFEEAYIPKEQKYSFLONPOTLSFSSESIP 60

Db 1 MFPTIPLSRLFDNAMLRAH; HQLAFTDYQEFEEAYIPKEQKYSFLONPOTLSFSSESIP 60

Oy 61 TPSNRETOQKSNLELLRISLLLIQSWLEPVQ 92

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Db 61 TPSNREETOOKSNLELLRISLLLIQSWLEPVQ 92

## RESULT 9

AAP91299  
ID AAP91299 standard; protein; 261 AA.

XX AC AAP91299;

XX DT 14-DEC-1989 (first entry).

XX DE Human nerve growth factor and human growth hormone fusion protein.

XX KW Human nerve growth factor; fusion protein; thrombin;

XX KW geriatric dementia; nervous disorders; human growth hormone.

XX OS Homo sapiens (human).

XX FH Key Location/Qualifiers

FT Region 1..140

FT Region 141..143

FT Region 144..261

XX PN EP329175-A.

XX PD 23-AUG-1989.

XX PF 17-FEB-1989; 89EP-0102795.

XX PR 19-FEB-1988; 88JP-0035042.

XX PA (TOYJ ) TOSOH CORP.

XX PI Ohtsuka E;

XX DR WPI; 1989-243092/34.

XX PT New human nerve growth factor gene encoding fusion protein

XX PT - having cleavage site for thrombin, useful for treating geriatric

XX PS Claim 36; page 31-32; 38pp; English.

XX CC Fusion protein consisting of human growth hormone at the  
N-terminal end (1st region), a 3 amino acid sequence representing  
thrombin recognition site, and human beta nerve growth factor (beta-NGF)  
at the C-terminal. Beta-NGF can be used to control geriatric dementia  
and other nervous disorders, and can be released from the fusion  
protein by incubation with thrombin (see AAN90577-8, AAP91034).

XX SQ Sequence 261 AA;

Query Match 58.3%; Score 465; DB 10; Length 261;

Best Local Similarity 98.9%; Pred. No. 6.8e-22;

Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MFPTIPLSLRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKEQKYSFLQNPTSLSFSESIP 60

DB 1 MFPTIPLSLRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKEQKYSFLQNPTSLSFSESIP 60

OY 61 TPSNREETOOKSNLELLRISLLLIQSWLEPVQ 92

DB 61 TPSNREETOOKSNLELLRISLLLIQSWLEPVQ 92

## RESULT 10

AAR11740  
ID AAR11740 standard; Protein; 262 AA.

XX AC AAR11740;

XX DT 25-MAR-2003 (updated)

DT 25-JUN-1991 (first entry)

XX DE Human growth hormone/human nerve growth factor beta fusion protein.

XX KW hGH; hNGF; nervous system diseases; dementia.

XX OS Homo sapiens.

XX PN JP03067598-A.

XX PD 22-MAR-1991.

XX PF 07-AUG-1989; 89JP-0202835.

XX PR 07-AUG-1989; 89JP-0202835.

XX PA (TOYJ ) TOSOH CORP.

XX DR WPI; 1991-128768/18.

XX DR N-PSDB; AA011578.

XX PT Purification of human neuron growth factor beta subunit-contg. protein  
by contacting with gel having cation exchange gp. in presence of  
urea

XX PS Disclosure ; fig 1; 7pp; Japanese.

XX CC A recombinant human nerve growth factor beta subunit-contg. protein  
can be produced as this fusion protein. It is purified by contacting  
a gel having a cation exchange gp. with the fusion protein, in the  
presence of urea. The purified protein is useful in a medicament  
for treating disorders of the nervous system, eg dementia.  
(Updated on 25-MAR-2003 to correct PF field.)

XX SQ Sequence 262 AA;

Query Match 58.3%; Score 465; DB 12; Length 262;

Best Local Similarity 98.9%; Pred. No. 6.8e-22;

Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MFPTIPLSLRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKEQKYSFLQNPTSLSFSESIP 60

DB 1 MFPTIPLSLRFDNAMLRAHRLHQLAFDTYOEFEEAYIPKEQKYSFLQNPTSLSFSESIP 60

OY 61 TPSNREETOOKSNLELLRISLLLIQSWLEPVQ 92

DB 61 TPSNREETOOKSNLELLRISLLLIQSWLEPVQ 92

## RESULT 11

AAR03255

ID AAR03255 standard; protein; 310 AA.

XX AC AAR03255;

XX DT 19-JUL-1990 (first entry)

XX DE Fusion protein of B-cell stimulatory factor-2 and B-cell

XX DE differentiation factor.

XX KW B-cell stimulatory factor-2; interleukin-6; B-cell differentiation;

XX KW interleukin-5; fusion protein.

XX OS Homo sapiens.

XX PN JP02013375-A.

XX PD 17-JAN-1990.

XX PF 01-JUL-1988; 88JP-0162556.

XX PR 01-JUL-1988; 88JP-0162556.

PA (TOYJ) TOSOH CORP.  
 XX  
 DR WPI: 1990-062207/09.  
 DR N-PSDB: AAQ02028.  
 XX  
 PT Prepn. of human B-cell differentiation factor - from specified DNA  
 PT sequence segment, by recombinant DNA technique, gives protein of  
 PT specified amino acid sequence.  
 XX  
 PS Claim 31; Page 9; 17pp; Japanese.  
 XX  
 CC The protein is produced by fusing DNA encoding hGF (IL-) with DNA  
 CC encoding bSF-2 (IL-5) and ligating the product into an expression vector  
 CC See also AAR05311 and AAR05313.  
 XX  
 SQ Sequence 310 AA;  
 Query Match 58.3%; Score 465; DB 11; Length 310;  
 Best Local Similarity 98.9%; Pred. No. 7.7e-22;  
 Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MFPTPLSRFLDNAMLRHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPOTSLSFSESIP 60  
 DB 1 MFPTPLSRFLDNAMLRHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPOTSLSFSESIP 60  
 QY 61 TPNREETOQKSNLELLRISILLIQSWLEPVQ 92  
 DB 61 TPNREETOQKSNLELLRISILLIQSWLEPVQ 92  
 RESULT 12  
 ABG31862  
 ID ABG31862 standard; Protein: 191 AA.  
 XX  
 AC ABG31862;  
 XX  
 DT 05-NOV-2002 (first entry)  
 XX  
 DE Mature human growth hormone (hGH), mutant #4.  
 XX  
 KW Human; growth hormone; hGH; Turner's syndrome; achondroplasia;  
 KW growth hormone deficiency in adults; GHDA; chronic renal insufficiency;  
 KW renal failure in children; acquired immune deficiency syndrome; AIDS;  
 KW AIDS wasting; cachexia; mutant; mutein.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 134  
 FT /note= "Wild type Arg substituted by Lys"  
 XX  
 PN W020025532-A2.  
 XX  
 PD 18-JUL-2002.  
 XX  
 PE 10-JAN-2002; 2002WO-DK00017.  
 XX  
 PR 11-JAN-2001; 2001DK-0000042.  
 PR 11-JAN-2001; 2001US-261411P.  
 XX  
 PA (MAXY-) MAXYGEN APS.  
 PA (MAXY-) MAXYGEN HOLDINGS LTD.  
 XX  
 PI Andersen KV, Drustup J, Christiansen J;  
 XX  
 DR WPI: 2002-608345/65.  
 XX  
 PT New conjugates exhibiting growth hormone activity, useful for treating  
 PT a disease or for manufacturing a medicament for treating a disease.  
 PT e.g. Turner's syndrome, growth hormone deficiency, achondroplasia, AIDS  
 PT wasting or cachexia -  
 XX

PS Claim 10; Page -: 74pp; English.  
 XX  
 CC The invention relates to new conjugates, which exhibit growth hormone  
 CC (GH) activity and comprise at least one non-polypeptide group covalently  
 CC attached to a GH polypeptide. The amino acid sequence of the conjugates  
 CC differs from that of wild type human GH in at least one introduced and at  
 CC least one removed amino acid residue comprising an attachment group for  
 CC the first non-polypeptide group. The conjugate or pharmaceutical  
 CC composition is useful for treating a disease or for manufacturing a  
 CC medicament for treating a disease, e.g. Turner's syndrome, GH deficiency  
 CC in adults (i.e. GHDA), achondroplasia, chronic renal insufficiency or  
 CC failure (including renal failure in children), acquired immune deficiency  
 CC syndrome (AIDS) wasting, cachexia in AIDS patients, or cachexia  
 CC associated with other diseases. The conjugates are useful for treating a  
 CC variety of disorders caused by growth hormone inadequacy. The present  
 CC sequence represents the amino acid sequence of a mutant human growth  
 CC hormone.  
 CC Note: The present sequence is not shown in the specification but is  
 CC derived from the wild type human growth hormone sequence given in SEQ ID  
 CC No.2 (see ABG31857).  
 XX  
 SQ Sequence 191 AA;  
 Query Match 58.2%; Score 464; DB 23; Length 191;  
 Best Local Similarity 71.0%; Pred. No. 6.1e-22;  
 Matches 103; Conservative 6; Mismatches 20; Indels 16; Gaps 4;  
 QY 2 FPTPLSRFLDNAMLRHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPOTSLSFSESIP 61  
 DB 1 FPTPLSRFLDNAMLRHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPOTSLSFSESIP 60  
 QY 62 PSNREETOQKSNLELLRISILLIQSWLEPVOLGCPRFVNOHLCGS-----HLVE 111  
 DB 61 PSNREETOQKSNLELLRISILLIQSWLEPVQF-LRSVFANSLSVYGASDSNVYDLKOLEE 119  
 QY 112 ALYLVCQ--ERGFFYTPKTRGIVEQ 134  
 DB 120 IQTLGRLSDG---SPKGTGIFKO 141  
 RESULT 13  
 ABG94860  
 ID ABG94860 standard; Protein: 141 AA.  
 XX  
 AC ABG94860;  
 XX  
 DT 03-DEC-2002 (first entry)  
 XX  
 DE Human growth hormone mutant hpt. (109-112).  
 XX  
 KW Growth hormone; placental lactogen; prolactin; active domain; hGH;  
 KW structure-function relationship; segment-substituted polypeptide;  
 KW mutant; mutein.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN US6428954-B1.  
 XX  
 PD 06-AUG-2002.  
 XX  
 PE 06-JUN-1995; 95US-0483039.  
 XX  
 PR 26-OCT-1989; 89US-0428066.  
 PR 27-APR-1992; 92US-0875204.  
 PR 13-OCT-1992; 92US-0960227.  
 PR 02-FEB-1994; 94US-0190723.  
 PR 28-OCT-1988; 88US-0264611.  
 XX  
 PA (GETH) GENENTECH INC.  
 PA Wells JA, Cunningham BC;  
 PI  
 XX

WP1: 2002-696875/75.

identifying active domains within cloned polypeptides of known amino acid sequence by substituting analog segments into the parent polypeptide is useful to determine the relationship between structure and function.

Example 1; Page -: 86pp; English.

The invention relates to identifying an unknown active domain in a region of known amino acid sequence in a parent polypeptide e.g. human growth hormone (hGH) which has been cloned and has a pre-identified biological activity, where the active domain interacts with a target when the parent polypeptide is in its native-folded form and the interaction is responsible for the biological activity comprising: (a) comparing the amino acid sequence or polypeptide structure in the region of known amino acid sequence of hGH with the amino acid sequence or polypeptide structure in a region of known amino acid sequence of an analogue polypeptide (e.g. prolactin, placental lactogen or porcine growth hormone) which has at least 15% homology with hGH alpha-carbon coordinates within about 2-3.5 angstroms of hGH alpha-carbon coordinates for about 60% of the analogue sequence where any interaction of the analogue with the target is different from target interaction with hGH; (b) substituting DNA encoding an analogous polypeptide segment from the analogue into DNA encoding the full length hGH, and expressing a segment-substituted polypeptide; (c) contacting the segment-substituted polypeptide with the target to determine interaction; (d) repeating steps (b) and (c) with a second analogous polypeptide segment; and (e) comparing the difference between activity of the first and second segment-substituted polypeptides as an indication of the location of the unknown active domain in hGH. The method is useful for determining the relationship between structure and function of known polypeptide sequences. The present sequence is that of human growth hormone mutant substituted with residues from an hGH analogue (prolactin, placental lactogen or porcine growth hormone).

Note: The present sequence is not shown in the specification but was created by the indexer using the mature hGH sequence and information contained in the specification.

Sequence 191 AA:

Synthetic.

US6428954-B1.

06-AUG-2002.

06-JUN-1995; 95US-0484039.

26-OCT-1989; 89US-0428066.

27-APR-1992; 92US-0875204.

13-OCT-1992; 92US-0960227.

02-FEB-1994; 94US-0190723.

28-OCT-1998; 88US-0264611.

(GETH ) GENENTECH INC.

Wells JA, Cunningham BC.

WPI: 2002-696875/75.

Identifying active domains within cloned polypeptides of known amino acid sequence by substituting analog segments into the parent polypeptide is useful to determine the relationship between structure and function.

Example 16; Page 7; 86pp; English.

The invention relates to identifying an unknown active domain in a region of known amino acid sequence in a parent polypeptide e.g. human growth hormone (hGH) which has been cloned and has a pre-identified biological activity, where the active domain interacts with a target when the parent polypeptide is in its native-folded form and the interaction is responsible for the biological activity comprising: (a) comparing the amino acid sequence or polypeptide structure in the region of known amino acid sequence of hGH with the amino acid sequence or polypeptide structure in a region of known amino acid sequence of an analogue polypeptide (e.g. prolactin, placental lactogen or porcine growth hormone) which has at least 15% homology with hGH alpha-carbon coordinates within about 2.5 angstroms of hGH alpha-carbon coordinates for about 60% of the analogue sequence, where any interaction of the analogue with the target is different from target interaction with hGH; (b) substituting DNA encoding an analogous polypeptide segment from the analogue into DNA encoding the full length hGH, and expressing a segment-substituted polypeptide; (c) contacting the segment-substituted polypeptide with the target to determine interaction; (d) repeating steps (b) and (c) with a second analogous polypeptide segment; and (e) comparing the difference between activity of the first and second segment-substituted polypeptides as an indication of the location of the unknown active domain in hGH. The method is useful for determining the relationship between structure and function of known polypeptide sequences. The present sequence is that of human growth hormone mutant substituted at functionally important residues and used in the method of the invention.

Note: The present sequence is not shown in the specification but was created by the indexer using the mature hGH sequence and information contained in the specification.

Search completed: September 16, 2003, 12:38:27  
Job time : 45.358 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 12:37:10 : Search time 23.93 seconds  
(without alignments)  
265,217 Million cell updates/sec

Title: us-09-423-100-7

Perfect score: 797

Sequence: 1 MFPTPLSLRFLDNAMLAHR.....IVQCCTSIKSLYQLENYCN 150

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA\*

- 1: /cgn2\_6/ptodata/1/1aa/5A-COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/1aa/5B-COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/1aa/6A-COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/1aa/6B-COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/1aa/PCTUS-COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score if the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	466	58.5	192	1	Sequence 1, Appl
2	461	57.8	191	4	Sequence 5, Appl
3	461	57.8	194	2	Sequence 4, Appl
4	461	57.8	194	3	Sequence 4, Appl
5	461	57.8	217	3	Sequence 10, Appl
6	461	57.8	217	3	Sequence 10, Appl
7	461	57.8	217	3	Sequence 10, Appl
8	461	57.8	217	3	Sequence 11, Appl
9	461	57.8	217	3	Sequence 1, Appl
10	461	57.8	241	4	Sequence 25, Appl
11	461	57.8	245	4	Sequence 66, Appl
12	461	57.8	274	3	Sequence 71, Appl
13	461	57.8	360	3	Sequence 73, Appl
14	455	57.1	191	4	Sequence 1, Appl
15	455	57.1	217	1	Sequence 4, Appl
16	455	57.1	217	1	Sequence 4, Appl
17	455	57.1	217	2	Sequence 51, Appl
18	455	57.1	217	2	Sequence 51, Appl
19	455	57.1	217	4	Sequence 4, Appl
20	454	57.0	400	4	Sequence 37, Appl
21	454	57.0	401	4	Sequence 36, Appl
22	448	56.2	191	3	Sequence 18, Appl
23	446	56.0	191	3	Sequence 16, Appl
24	446	56.0	191	3	Sequence 20, Appl
25	365.5	45.9	176	3	Sequence 1, Appl
26	365.5	45.9	176	4	Sequence 1, Appl
27	359.5	45.1	176	3	Sequence 2, Appl

Sequence 2, Appl  
Patent No. 5424199  
Sequence 5, Appl  
Sequence 5, Appl  
Sequence 5, Appl  
Sequence 8, Appl  
Sequence 39, Appl  
Sequence 39, Appl  
Sequence 48, Appl  
Sequence 48, Appl  
Sequence 45, Appl  
Sequence 45, Appl  
Sequence 8, Appl  
Sequence 1, Appl  
Sequence 4, Appl  
Sequence 6, Appl  
Sequence 2, Appl

ALIGNMENTS

RESULT 1  
US-08-093-383-1  
Sequence 1, Application US/08093383  
Patent No. 5489529  
GENERAL INFORMATION:  
APPLICANT: DeBoer, Herman A.  
APPLICANT: Heyneker, Herbert L.  
APPLICANT: Seeburg, Peter H.  
TITLE OF INVENTION: DNA for Expression of Bovine Growth Hormone  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESS: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patlin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/093-383  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/619827  
FILING DATE: 28-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/198824  
FILING DATE: 05-APR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 06/632361  
FILING DATE: 19-JUL-1984  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 06/303687  
FILING DATE: 18-SEP-1981  
ATTORNEY/AGENT INFORMATION:  
NAME: Johnston, Sean A.  
REGISTRATION NUMBER: P35,910  
REFERENCE/DOCKET NUMBER: 46C4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-3562  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 192 amino acids  
TYPE: amino acid



```

; TOPOLOGY: linear
US-08-093-383-1

Query Match      58.5%  Score 466;  DB 1;  Length 192;
Best Local Similarity 70.3%  Pred. No. 1.1e-42;
Matches 103;  Conservative 7;  Mismatches 20;  Indels 16;  Gaps 4;

QY 1 MFPTPLSRFLDNAMLAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPOTSLSFSESIP 60
DB 1 MFPTPLSRFLDNAMLAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPOTSLSFSESIP 60
QY 61 TPSNREETQOKSNLELLRISLLLIQSWLEPVOLGTGPRFVNOHLCGS-----HLV 110
DB 61 TPSNREETQOKSNLELLRISLLLIQSWLEPVQF-LRSVFANSVLVYGASDSNVYDLKDL 119

QY 111 EALYLVCG--ERGFYTPKTRGIVEQ 134
DB 120 EGIOTLMGRLEDG---SPRTGQIFKQ 142

RESULT 2
US-09-284-878-5
; Sequence 5, Application US/09284878
; Patent No. 6342375
; GENERAL INFORMATION:
; APPLICANT: Olazaran, Martha Guerrero
; APPLICANT: Saldana, Hugo Barrera
; APPLICANT: Salvado, Jose Maria Viader
; TITLE OF INVENTION: Genetically Modified Methylotrophic P. pastoris Yeast for the
; TITLE OF INVENTION: Production and Secretion of the Human Growth Hormone
; FILE REFERENCE: 1829.0010000
; CURRENT APPLICATION NUMBER: US/09/284.878
; PRIOR FILING DATE: 1999-07-21
; PRIOR APPLICATION NUMBER: PCT/MX97/00033
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-284-878-5

Query Match      57.8%  Score 461;  DB 4;  Length 191;
Best Local Similarity 70.3%  Pred. No. 3.7e-42;
Matches 102;  Conservative 7;  Mismatches 20;  Indels 16;  Gaps 4;

QY 2 FPTPLSRFLDNAMLAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPOTSLSFSESIP 61
DB 1 FPTPLSRFLDNAMLAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPOTSLSFSESIP 60
QY 62 PSNREETQOKSNLELLRISLLLIQSWLEPVOLGTGPRFVNOHLCGS-----HLVE 111
DB 61 PSNREETQOKSNLELLRISLLLIQSWLEPVQF-LRSVFANSVLVYGASDSNVYDLKDL 119

QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134
DB 120 GIOTLMGRLEDG---SPRTGQIFKQ 141

RESULT 3
US-08-383-621-4
; Sequence 4, Application US/08383621
; Patent No. 5951972
; GENERAL INFORMATION:
; APPLICANT: Daley, Michael J.
; APPLICANT: Buckwalter, Brian L.
; APPLICANT: Cady, Susan M.
; APPLICANT: Shieh, Hong-Ming
; APPLICANT: Bohlen, Peter
; APPLICANT: Seddon, Andrew P.
; TITLE OF INVENTION: Stabilization of Somatotropins And Other
; TITLE OF INVENTION: Proteins By Modification Of Cysteine Residues

```

```

; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. Estelle J. Tsevdos
; STREET: 1937 West Main Street, P.O. Box 60
; CITY: Stamford
; STATE: Connecticut
; COUNTRY: U.S.A.
; ZIP: 06904-0060
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/383.621
; FILING DATE: 06-FEB-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/766,142
; FILING DATE: 25-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsevdos, Estelle J.
; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,278-91
; TELEPHONE: 203-321-2756
; TELEFAX: 203-321-2971
; TELEX: 203-710-474-4059
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 194 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-383-621-4

Query Match      57.8%  Score 461;  DB 2;  Length 194;
Best Local Similarity 70.3%  Pred. No. 3.8e-42;
Matches 102;  Conservative 7;  Mismatches 20;  Indels 16;  Gaps 4;

QY 2 FPTPLSRFLDNAMLAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPOTSLSFSESIP 61
DB 4 FPTPLSRFLDNAMLAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPOTSLSFSESIP 63
QY 62 PSNREETQOKSNLELLRISLLLIQSWLEPVOLGTGPRFVNOHLCGS-----HLVE 111
DB 54 PSNREETQOKSNLELLRISLLLIQSWLEPVQF-LRSVFANSVLVYGASDSNVYDLKDL 122

QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134
DB 123 GIOTLMGRLEDG---SPRTGQIFKQ 144

RESULT 4
US-08-459-906-4
; Sequence 4, Application US/08459906
; Patent No. 6010995
; GENERAL INFORMATION:
; APPLICANT: Daley, Michael J.
; APPLICANT: Buckwalter, Brian L.
; APPLICANT: Cady, Susan M.
; APPLICANT: Shieh, Hong-Ming
; APPLICANT: Bohlen, Peter
; APPLICANT: Seddon, Andrew P.
; TITLE OF INVENTION: Stabilization of Somatotropins and Other
; TITLE OF INVENTION: Proteins By Modification of Cysteine Residues
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: U.S.A.

```





```
; NAME/KEY: peptide
; LOCATION: 133..153
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 192..210
; OTHER INFORMATION: /note- "The peptides above are
; OTHER INFORMATION: depicted in Figure 1"
US-08-759-628-11

Query Match          57.8%; Score 461; DB 3; Length 217;
Best Local Similarity 70.3%; Pred. No. 4.4e-42;
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;

QY 2 FPTPLSLRFDNAMLRAHRLHQLAFDTYQEEFAEYIPKEQKYSFLQNPTSLSFSESIPT 61
DB 27 FPTPLSLRFDNAMLRAHRLHQLAFDTYQEEFAEYIPKEQKYSFLQNPTSLSFSESIPT 86
QY 62 PSNREETOQKSNLELLRLISLLLIQSWLEPVLGTGPRFVNOHLGCS-----HLVE 111
DB 87 PSNREETOQKSNLELLRLISLLLIQSWLEPVLGTGPRFVNOHLGCS-----HLVE 145
QY 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134
DB 146 GIOTLMGRLEDG---SPRTGOIFKQ 167

RESULT 9
US-09-284-878-1
; Sequence 1, Application US/09284878
; Patent No. 6342375
; GENERAL INFORMATION:
; APPLICANT: Olazaran, Martha Guerrero
; APPLICANT: Saldana, Hugo Barrera
; APPLICANT: Salgado, Jose Maria Viader
; TITLE OF INVENTION: Genetically Modified Methylophilic P. pastoris Yeast for the
; TITLE OF INVENTION: Production and Secretion of the Human Growth Hormone
; FILE REFERENCE: 1829.0010000
; CURRENT APPLICATION NUMBER: US/09/284,878
; PRIOR FILING DATE: 1999-07-21
; PRIOR APPLICATION NUMBER: PCT/MX97/00033
; PRIOR FILING DATE: 1997-10-24
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-284-878-1

Query Match          57.8%; Score 461; DB 4; Length 217;
Best Local Similarity 70.3%; Pred. No. 4.4e-42;
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;

QY 2 FPTPLSLRFDNAMLRAHRLHQLAFDTYQEEFAEYIPKEQKYSFLQNPTSLSFSESIPT 61
DB 27 FPTPLSLRFDNAMLRAHRLHQLAFDTYQEEFAEYIPKEQKYSFLQNPTSLSFSESIPT 86
QY 62 PSNREETOQKSNLELLRLISLLLIQSWLEPVLGTGPRFVNOHLGCS-----HLVE 111
DB 87 PSNREETOQKSNLELLRLISLLLIQSWLEPVLGTGPRFVNOHLGCS-----HLVE 145
QY 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134
DB 146 GIOTLMGRLEDG---SPRTGOIFKQ 167

RESULT 10
US-09-424-620B-25
; Sequence 25, Application US/09424620B
; Patent No. 6391585
; GENERAL INFORMATION:
; APPLICANT: HANIL SYNTHETIC FIBER CO., LTD.
; APPLICANT: JANG, Ki-Ryong
```

```
; MOON, Jae-Woor
; BAE, Choon-Soo
; YANG, Doo-Suk
; LEE, Jee-Won
; SEONG, Baik-Lin
; TITLE OF INVENTION: Process for preparing recombinant proteins using highly
; efficient expression vector from Saccharomyces cerevisiae
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BACHMAN & LAPORTE, P.C.
; STREET: Suite 1201, 910 Chapel Street
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: U.S.A.
; ZIP: 06510-2802
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disket; 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM
; OPERATING SYSTEM: Windows 95/98
; SOFTWARE: MS WORD
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/424,620B
; FILING DATE: 24-Nov-6391585-1999
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 241 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: PROTEIN
; SEQUENCE DESCRIPTION: SEQ ID NO: 25:
US-09-424-620B-25

Query Match          57.8%; Score 461; DB 4; Length 241;
Best Local Similarity 70.3%; Pred. No. 9e-42;
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;

QY 2 FPTPLSLRFDNAMLRAHRLHQLAFDTYQEEFAEYIPKEQKYSFLQNPTSLSFSESIPT 61
DB 51 FPTPLSLRFDNAMLRAHRLHQLAFDTYQEEFAEYIPKEQKYSFLQNPTSLSFSESIPT 110
QY 62 PSNREETOQKSNLELLRLISLLLIQSWLEPVLGTGPRFVNOHLGCS-----HLVE 111
DB 111 PSNREETOQKSNLELLRLISLLLIQSWLEPVLGTGPRFVNOHLGCS-----HLVE 169
QY 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134
DB 170 GIOTLMGRLEDG---SPRTGOIFKQ 191

RESULT 11
US-09-280-030-66
; Sequence 66, Application US/09280030A
; Patent No. 6506595
; GENERAL INFORMATION:
; APPLICANT: Sato, Seiji
; APPLICANT: Higashikuni, Naohiko
; APPLICANT: Kudo, Toshiyuki
; APPLICANT: Kondo, Masaki
; TITLE OF INVENTION: DNAS ENCODING NEW FUSION PROTEINS AND PROCESSES FOR
; TITLE OF INVENTION: PREPARING USEFUL POLYPEPTIDES THROUGH EXPRESSION OF THE
; FILE REFERENCE: 382.1026
; CURRENT APPLICATION NUMBER: US/09/280,030A
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: J10-87339/1998
; EARLIER FILING DATE: 1998-03-11
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 66
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
```

;  
US-08-784-582-71

Query Match	57.88;	Score 461;	DB 3;	Length 360;
Best Local Similarity	70.30;	Pred. No. 8.5e-42;		

```
Matches 102; Conservative 7; Mismatch 20; Indels 16; Gaps 4;
QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQEEFAYIPKEOKYSFLONPQTSLSFSSEIPT 61
Db 27 FPTPLSRFDNAMLRAHRLHQLAFDTYQEEFAYIPKEOKYSFLONPQTSLSFSSEIPT 86
QY 62 PSNREETQOKSNLELLRISLLLSQSWLEPVQGTGPRFVNOHLGGS-----HLVE 111
Db 87 PSNREETQOKSNLELLRISLLLSQSWLEPVQF-LRSVFANSLVYGSASNSVYDLKDLFE 145
QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134
Db 146 GIQTLGRLDGG---SPRTGQIFKQ 167

RESULT 14
US-09-465-461-1
; Sequence 1, Application US/09465461
; Patent No. 6348444
; GENERAL INFORMATION:
; APPLICANT: C APPEL, Scott
; TITLE OF INVENTION: Human Growth Hormone to stimulate hematopoiesis and immune reconst
; TITLE OF INVENTION: after hematopoietic stem cell transplantation in humans
; FILE REFERENCE: CHAPPEL-6.1
; CURRENT APPLICATION NUMBER: US/09/465.461
; PRIOR FILING DATE: 1999-12-17
; PRIOR APPLICATION NUMBER: 60/112,668
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 191
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-465-461-1

Query Match 57.1%; Score 455; DB 1; Length 191;
Best Local Similarity 69.7%; Pred. No. 1.7e-41;
Matches 101; Conservative 7; Mismatches 21; Indels 16; Gaps 4;
QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQEEFAYIPKEOKYSFLONPQTSLSFSSEIPT 61
Db 1 FPTPLSRFDNAMLRAHRLHQLAFDTYQEEFAYIPKEOKYSFLONPQTSLSFSSEIPT 60
QY 62 PSNREETQOKSNLELLRISLLLSQSWLEPVQGTGPRFVNOHLGGS-----HLVE 111
Db 61 PSNREETQOKSNLELLRISLLLSQSWLEPVQF-LRSVFANSLVYGSASNSVYDLKDLFE 179
QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134
Db 120 GIQTLGRLDGG---SPRTGQIFKQ 141

RESULT 15
US-08-187-756C-4
; Sequence 4, Application US/08187756C
; Patent No. 5597709
; GENERAL INFORMATION:
; APPLICANT: ROSEN, ET AL.
; TITLE OF INVENTION: Human Growth Hormone
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
```

```
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/187.756C
FILING DATE: January 27, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,114
REFERENCE/DOCKET NUMBER: 325860-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 217 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-187-756C-4

Query Match 57.1%; Score 455; DB 1; Length 217;
Best Local Similarity 69.7%; Pred. No. 2e-41;
Matches 101; Conservative 7; Mismatches 21; Indels 16; Gaps 4;
QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQEEFAYIPKEOKYSFLONPQTSLSFSSEIPT 61
Db 27 FPTPLSRFDNAMLRAHRLHQLAFDTYQEEFAYIPKEOKYSFLONPQTSLSFSSEIPT 86
QY 62 PSNREETQOKSNLELLRISLLLSQSWLEPVQGTGPRFVNOHLGGS-----HLVE 111
Db 87 PSNREETQOKSNLELLRISLLLSQSWLEPVQF-LRSVFANSLVYGSASNSVYDLKDLFE 145
QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134
Db 146 GIQTLGRLDGG---SPRTGQIFKQ 167
```

Search completed: September 16, 2003, 12:41:25  
Job time : 24.93 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 12:38:55 : Search time 47.2763 seconds  
(without alignments)  
472.415 Million cell updates/sec

Title: US-09-423-100-7

Perfect score: 797

Sequence: 1 MFPTPLSLRFLDNAMLAHR.....IVROCTSLCSLYOLENYCN 150

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 556269 seqs, 14893369 residues

Total number of hits satisfying chosen parameters: 556269

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	797	100.0	150	14	US-10-054-873-7
2	555.5	69.7	107	14	US-10-054-873-6
3	470	59.0	92	14	US-10-054-873-2
4	461	57.8	191	11	US-09-984-010-23
5	461	57.8	191	12	US-10-153-207-1
6	461	57.8	191	12	US-10-400-377-1
7	461	57.8	191	12	US-10-400-708-1
8	461	57.8	191	12	US-10-298-148-1
9	461	57.8	214	12	US-10-153-207-6
10	461	57.8	217	9	US-09-929-918-9
11	461	57.8	245	9	US-09-380-030-66
12	455	57.1	217	9	US-09-853-688-2
13	455	57.1	217	12	US-09-969-748C-4
14	454	57.0	217	10	US-09-804-409A-16
15	448	56.2	217	9	US-09-853-688-4

16	445	55.8	191	12	US-09-824-200-12	Sequence 12, Appl
17	440	55.2	191	12	US-10-300-822-1	Sequence 1, Appl
18	433	54.3	191	12	US-10-289-845-3	Sequence 3, Appl
19	399	50.1	217	9	US-09-850-887-3	Sequence 3, Appl
20	381	47.8	163	15	US-10-043-487-350	Sequence 350, App
21	381	47.8	191	12	US-10-153-207-2	Sequence 2, Appl
22	373	46.8	229	15	US-10-103-113-411	Sequence 411, App
23	345	43.3	246	15	US-10-188-246-18	Sequence 18, Appl
24	301.5	37.8	190	12	US-10-153-207-3	Sequence 3, Appl
25	294.5	37.0	138	10	US-09-861-687-19	Sequence 19, Appl
26	294	36.9	52	14	US-10-054-873-5	Sequence 5, Appl
27	286	35.9	147	9	US-09-736-611-8	Sequence 8, Appl
28	286	35.9	147	9	US-09-740-359-7	Sequence 7, Appl
29	284.5	35.7	124	9	US-09-894-711-18	Sequence 18, Appl
30	284.5	35.7	144	9	US-09-736-611-6	Sequence 6, Appl
31	284.5	35.7	144	9	US-09-740-359-5	Sequence 5, Appl
32	284.5	35.7	146	10	US-09-894-711-5	Sequence 5, Appl
33	278.5	34.9	51	11	US-09-858-935A-5	Sequence 5, Appl
34	278.5	34.9	51	14	US-10-028-410-3	Sequence 3, Appl
35	278	34.9	117	9	US-09-280-030-63	Sequence 63, Appl
36	277	34.8	96	10	US-09-547-563-4	Sequence 4, Appl
37	275.5	34.6	124	9	US-09-736-611-12	Sequence 12, Appl
38	275.5	34.6	124	9	US-09-740-359-12	Sequence 12, Appl
39	275.5	34.6	124	10	US-09-894-711-12	Sequence 12, Appl
40	275.5	34.6	125	9	US-09-736-611-10	Sequence 10, Appl
41	275.5	34.6	125	9	US-09-740-359-10	Sequence 10, Appl
42	275.5	34.6	125	10	US-09-894-711-10	Sequence 10, Appl
43	273	34.3	50	14	US-10-066-009A-3	Sequence 3, Appl
44	271	34.0	96	10	US-09-547-563-5	Sequence 5, Appl
45	270	33.9	110	9	US-09-205-658-125	Sequence 125, App

ALIGNMENTS

RESULT 1

US-10-054-873-7  
: Sequence 7, Application US/10054873  
: Publication No. US20020164712A1  
: GENERAL INFORMATION:  
: APPLICANT: Gan, Zhong Ru  
: TITLE OF INVENTION: Chimeric Protein Containing an  
: Intramolecular Chapterone-Like Sequence  
: NUMBER OF SEQUENCES: 7  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Townsend and Townsend and Crew LLP  
: STREET: Two Embarcadero Center, Eighth Floor  
: CITY: San Francisco  
: STATE: California  
: COUNTRY: USA  
: ZIP: 94111-3834  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/10/054,873  
: FILING DATE: 22-Jan-2002  
: CLASSIFICATION: <Unknown>  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: WO PCT/CN98/00052  
: FILING DATE: 31-MAR-1998  
: APPLICATION NUMBER: US 09/423,100  
: FILING DATE: 11-DEC-2000  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Mycroft, Frank J  
: REGISTRATION NUMBER: 46,946  
: REFERENCE/DOCKET NUMBER: 020167-00013005  
: INFORMATION FOR SEQ ID NO: 7:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 150 amino acids  
: TYPE: amino acid

```

; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-10-054-873-7

Query Match          100.0%; Score 797; DB 14; Length 150;
Best Local Similarity 100.0%; Pred. No. 1.4e-80;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFPTPLSLFDNAMLRAHRLHQLAFDTYQEEFAYIPKEQKYSFLQNPQTSLSFSESIP 60
DB 1 MFPTPLSLFDNAMLRAHRLHQLAFDTYQEEFAYIPKEQKYSFLQNPQTSLSFSESIP 60
QY 61 TPSNREETQOKSNLELLRISLLLIQSWLEPVOLGTGPRFVNOHLCGSHLVEALYLVCGER 120
DB 61 TPSNREETQOKSNLELLRISLLLIQSWLEPVOLGTGPRFVNOHLCGSHLVEALYLVCGER 120
QY 121 GFFYTPKTRGIVEQCCTSIQSLYQLENYCN 150
DB 121 GFFYTPKTRGIVEQCCTSIQSLYQLENYCN 150

RESULT 2
US-10-054-873-6
; Sequence 6, Application US/10054873
; Publication No. US20020164712A1
; GENERAL INFORMATION:
; APPLICANT: Gan, Zhong Ru
; TITLE OF INVENTION: Chimeric Protein Containing an
; Intramolecular Chaperone-Like Sequence
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/054,873
; FILING DATE: 22-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/CN98/00052
; FILING DATE: 31-MAR-1998
; APPLICATION NUMBER: 39/423,100
; FILING DATE: 11-DEC-2000
; ATTORNEY/AGENT INFORMATION:
; NAME: Mycroft, Frank J.
; REGISTRATION NUMBER: 46,946
; REFERENCE/DOCKET NUMBER: 020167-000130US
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 107 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-054-873-6

Query Match          69.7%; Score 555.5; DB 14; Length 107;
Best Local Similarity 71.3%; Pred. No. 5.9e-54;
Matches 107; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MFPTPLSLFDNAMLRAHRLHQLAFDTYQEEFAYIPKEQKYSFLQNPQTSLSFSESIP 60
DB 1 MFPTPLSLFDNAMLRAHRLHQLAFDTYQEEFAYIPKEQKYSFLQNPQTSLSFSESIP 60
QY 61 TPSNREETQOKSNLELLRISLLLIQSWLEPVOLGTGPRFVNOHLCGSHLVEALYLVCGER 120
DB 61 TPSNREETQOKSNLELLRISLLLIQSWLEPVOLGTGPRFVNOHLCGSHLVEALYLVCGER 120
QY 121 GFFYTPKTRGIVEQCCTSIQSLYQLENYCN 150
DB 121 GFFYTPKTRGIVEQCCTSIQSLYQLENYCN 150

RESULT 3
US-10-054-873-2
; Sequence 2, Application US/10054873
; Publication No. US20020164712A1
; GENERAL INFORMATION:
; APPLICANT: Gan, Zhong Ru
; TITLE OF INVENTION: Chimeric Protein Containing an
; Intramolecular Chaperone-Like Sequence
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/054,873
; FILING DATE: 22-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/CN98/00052
; FILING DATE: 31-MAR-1998
; APPLICATION NUMBER: 39/423,100
; FILING DATE: 11-DEC-2000
; ATTORNEY/AGENT INFORMATION:
; NAME: Mycroft, Frank J.
; REGISTRATION NUMBER: 46,946
; REFERENCE/DOCKET NUMBER: 020167-000130US
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 92 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-054-873-2

Query Match          59.0%; Score 470; DB 14; Length 92;
Best Local Similarity 100.0%; Pred. No. 1.5e-44;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFPTPLSLFDNAMLRAHRLHQLAFDTYQEEFAYIPKEQKYSFLQNPQTSLSFSESIP 60
DB 1 MFPTPLSLFDNAMLRAHRLHQLAFDTYQEEFAYIPKEQKYSFLQNPQTSLSFSESIP 60
QY 61 TPSNREETQOKSNLELLRISLLLIQSWLEPVOLGTGPRFVNOHLCGSHLVEALYLVCGER 120
DB 61 TPSNREETQOKSNLELLRISLLLIQSWLEPVOLGTGPRFVNOHLCGSHLVEALYLVCGER 120
QY 121 GFFYTPKTRGIVEQCCTSIQSLYQLENYCN 150
DB 121 GFFYTPKTRGIVEQCCTSIQSLYQLENYCN 150

RESULT 4
US-09-984-010-23
; Sequence 23, Application US/09984010
; Publication No. US20030104578A1
; GENERAL INFORMATION:

```



APPLICANT: Ballance, David James  
TITLE OF INVENTION: RECOMBINANT FUSION PROTEINS TO GROWTH HORMONE  
AND SERUM ALBUMIN  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, L.L.P.  
STREET: 1300 I Street, NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/984,010  
FILING DATE: 21-May-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/091,873  
FILING DATE: 25-JUN-1998  
APPLICATION NUMBER: PCT/GB96/03164  
FILING DATE: 19-DEC-1996  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 191 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 23:  
US-09-984-010-23

Query Match 57.8% Score 461 DB 11 Length 191  
Best Local Similarity 70.3% Pred. No. 3.9e-43  
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;  
QY 2 FPTPLSLRFDNMLRAHRLHQLAFDTYQEEAYIPKEQKYSFLONPQTSLSFSSEIPT 61  
Db 1 FPTPLSLRFDNMLRAHRLHQLAFDTYQEEAYIPKEQKYSFLONPQTSLSFSSEIPT 60  
QY 62 PSNREETQOKSNLELLRLISLLIQSWLEPVQGTGPRFVNQHLGCS-----HLVE 111  
Db 61 PSNREETQOKSNLELLRLISLLIQSWLEPVQGTGPRFVNQHLGCS-----HLVE 111  
QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134  
Db 120 GIQTLGRLGLEDG---SPRTGQIFKQ 141

RESULT 5  
US-10-153-207-1  
Sequence 1, Application US/10153207  
Publication No. US20030153003A1  
GENERAL INFORMATION:  
APPLICANT: James A. Wells  
TITLE OF INVENTION: GROWTH HORMONE VARIANTS  
FILE REFERENCE: 669.12-US-C7  
CURRENT APPLICATION NUMBER: US/10/153,207  
PRIOR FILING DATE: 2002-05-22  
PRIOR APPLICATION NUMBER: 08/479,884  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 08/190,723  
PRIOR FILING DATE: 1994-02-02  
PRIOR APPLICATION NUMBER: 07/960,227  
PRIOR FILING DATE: 1992-10-13  
PRIOR APPLICATION NUMBER: 07/875,204  
PRIOR FILING DATE: 1992-04-27  
PRIOR APPLICATION NUMBER: 07/428,066

PRIOR FILING DATE: 1989-10-26  
PRIOR APPLICATION NUMBER: 07/264,611  
PRIOR FILING DATE: 1988-10-28  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 191  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-10-153-207-1  
Query Match 57.8% Score 461 DB 12 Length 191  
Best Local Similarity 70.3% Pred. No. 3.9e-43  
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;  
QY 2 FPTPLSLRFDNMLRAHRLHQLAFDTYQEEAYIPKEQKYSFLONPQTSLSFSSEIPT 61  
Db 1 FPTPLSLRFDNMLRAHRLHQLAFDTYQEEAYIPKEQKYSFLONPQTSLSFSSEIPT 60  
QY 62 PSNREETQOKSNLELLRLISLLIQSWLEPVQGTGPRFVNQHLGCS-----HLVE 111  
Db 61 PSNREETQOKSNLELLRLISLLIQSWLEPVQGTGPRFVNQHLGCS-----HLVE 111  
QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134  
Db 120 GIQTLGRLGLEDG---SPRTGQIFKQ 141

RESULT 6  
US-10-400-377-1  
Sequence 1, Application US/1040,377  
Publication No. US20030162949A1  
GENERAL INFORMATION:  
APPLICANT: Cox III, George N  
TITLE OF INVENTION: Derivatives of Growth Hormone and Related Proteins  
FILE REFERENCE: 4152-1-PUS  
CURRENT APPLICATION NUMBER: US/10/400,377  
PRIOR FILING DATE: 2003-03-26  
PRIOR APPLICATION NUMBER: US/09/462,941  
PRIOR FILING DATE: 2000-01-14  
PRIOR APPLICATION NUMBER: 60/052,516  
PRIOR FILING DATE: 1997-07-14  
NUMBER OF SEQ ID NOS: 41  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 191  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-10-400-377-1

Query Match 57.8% Score 461 DB 12 Length 191  
Best Local Similarity 70.3% Pred. No. 3.9e-43  
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;  
QY 2 FPTPLSLRFDNMLRAHRLHQLAFDTYQEEAYIPKEQKYSFLONPQTSLSFSSEIPT 61  
Db 1 FPTPLSLRFDNMLRAHRLHQLAFDTYQEEAYIPKEQKYSFLONPQTSLSFSSEIPT 60  
QY 62 PSNREETQOKSNLELLRLISLLIQSWLEPVQGTGPRFVNQHLGCS-----HLVE 111  
Db 61 PSNREETQOKSNLELLRLISLLIQSWLEPVQGTGPRFVNQHLGCS-----HLVE 111  
QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134  
Db 120 GIQTLGRLGLEDG---SPRTGQIFKQ 141

RESULT 7  
US-10-400-708-1  
Sequence 1, Application US/10400708  
Publication No. US20030166865A1  
GENERAL INFORMATION:

APPLICANT: Cox III, George N  
APPLICANT: Bolder Biotechnology, Inc.  
TITLE OF INVENTION: Derivatives of Growth Hormone and Related Proteins  
FILE REFERENCE: 4152-1-PUS  
CURRENT APPLICATION NUMBER: US/10/400,708  
CURRENT FILING DATE: 2003-03-26  
PRIOR APPLICATION NUMBER: US/09/462,941  
PRIOR FILING DATE: 2000-01-14  
PRIOR APPLICATION NUMBER: 60/052,516  
PRIOR FILING DATE: 1997-07-14  
NUMBER OF SEQ ID NOS: 41  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 1  
LENGTH: 191  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-400-708-1

Query Match 57.8% Score 461; DB 12; Length 191;  
Best Local Similarity 70.3%; Pred. No. 3.9e-43;  
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSESPT 61  
DB 1 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSESPT 60  
QY 62 PSNREETQOKSNLELLRISLLIQSWLEPVLQGTGPRFVNHQCGS-----HLVE 111  
DB 61 PSNREETQOKSNLELLRISLLIQSWLEPVLQGTGPRFVNHQCGS-----HLVE 110  
QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134  
DB 120 GIQTLMGRLEDG---SPRTQIQFKQ 141

RESULT 8  
US-10-298-148-1  
Sequence 1, Application US/10298148  
Publication No. US20030171284A1  
GENERAL INFORMATION:  
APPLICANT: Cox III, George N  
APPLICANT: Bolder Biotechnology, Inc.  
TITLE OF INVENTION: Derivatives of Growth Hormone and Related Proteins  
FILE REFERENCE: 4152-1-PUS  
CURRENT APPLICATION NUMBER: US/10/298,148  
CURRENT FILING DATE: 2002-11-15  
PRIOR APPLICATION NUMBER: US/09/462,941  
PRIOR FILING DATE: 2000-01-14  
PRIOR APPLICATION NUMBER: 60/052,516  
PRIOR FILING DATE: 1997-07-14  
NUMBER OF SEQ ID NOS: 41  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 1  
LENGTH: 191  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-298-148-1

Query Match 57.8% Score 461; DB 12; Length 191;  
Best Local Similarity 70.3%; Pred. No. 3.9e-43;  
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSESPT 61  
DB 1 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSESPT 60  
QY 62 PSNREETQOKSNLELLRISLLIQSWLEPVLQGTGPRFVNHQCGS-----HLVE 111  
DB 61 PSNREETQOKSNLELLRISLLIQSWLEPVLQGTGPRFVNHQCGS-----HLVE 110  
QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134  
DB 120 GIQTLMGRLEDG---SPRTQIQFKQ 141

RESULT 9  
US-10-153-207-6  
Sequence 6, Application US/10153207  
Publication No. US20030153003A1  
GENERAL INFORMATION:  
APPLICANT: James A. Wells  
APPLICANT: Brian C. Cunningham  
TITLE OF INVENTION: GROWTH HORMONE VARIANTS  
FILE REFERENCE: 659.12-US-C7  
CURRENT APPLICATION NUMBER: US/10/153,207  
CURRENT FILING DATE: 2002-05-22  
PRIOR APPLICATION NUMBER: 08/479,884  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 08/190,723  
PRIOR FILING DATE: 1994-02-02  
PRIOR APPLICATION NUMBER: 07/960,227  
PRIOR FILING DATE: 1992-10-13  
PRIOR APPLICATION NUMBER: 07/875,264  
PRIOR FILING DATE: 1992-04-27  
PRIOR APPLICATION NUMBER: 07/428,066  
PRIOR FILING DATE: 1989-10-26  
PRIOR APPLICATION NUMBER: 07/264,611  
PRIOR FILING DATE: 1986-10-28  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6  
LENGTH: 214  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-153-207-6

Query Match 57.8% Score 461; DB 12; Length 214;  
Best Local Similarity 70.3%; Pred. No. 4.5e-43;  
Matches 102; Conservative 7; Mismatches 20; Indels 15; Gaps 4;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSESPT 61  
DB 24 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSESPT 83  
QY 62 PSNREETQOKSNLELLRISLLIQSWLEPVLQGTGPRFVNHQCGS-----HLVE 111  
DB 84 PSNREETQOKSNLELLRISLLIQSWLEPVLQGTGPRFVNHQCGS-----HLVE 142  
QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134  
DB 140 GIQTLMGRLEDG---SPRTQIQFKQ 164

RESULT 10  
US-09-929-918-9  
Sequence 9, Application US/09929918  
Patent No. US20020090678A1  
GENERAL INFORMATION:  
APPLICANT: Kordyum, Vitaliy A.  
APPLICANT: Chernykh, Svillana I.  
APPLICANT: Slavchenko, Iryna Yu.  
APPLICANT: Vozianov, Oleksandr  
TITLE OF INVENTION: PHAGE-DEPENDENT SUPER PRODUCTION OF  
TITLE OF INVENTION: BIOLOGICALLY ACTIVE PROTEIN AND PEPTIDES  
FILE REFERENCE: PHAGE.096A  
CURRENT APPLICATION NUMBER: US/09/929,918  
CURRENT FILING DATE: 2001-08-15  
PRIOR APPLICATION NUMBER: 09/318,288  
PRIOR FILING DATE: 1999-05-25  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9  
LENGTH: 217  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-929-918-9

Query Match 57.8% Score 461; DB 12; Length 191;  
Best Local Similarity 70.3%; Pred. No. 3.9e-43;  
Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;

QY 2 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSESPT 61  
DB 1 FPTPLSRFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSESPT 60  
QY 62 PSNREETQOKSNLELLRISLLIQSWLEPVLQGTGPRFVNHQCGS-----HLVE 111  
DB 61 PSNREETQOKSNLELLRISLLIQSWLEPVLQGTGPRFVNHQCGS-----HLVE 110  
QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134  
DB 120 GIQTLMGRLEDG---SPRTQIQFKQ 141

```

: : : : :
: : FILE REFERENCE: WCM78
: : CURRENT APPLICATION NUMBER: US/09/853,688
: : CURRENT FILING DATE: 2001-05-14
: : NUMBER OF SEQ ID NOS: 66
: : SOFTWARE: PatentIn Ver. 2.1
: : SEQ ID NO. 2
: : LENGTH: 217
: : TYPE: PRT
: : ORGANISM: Homo sapiens
: : US-09-853,688-2

```

DB 27 FPTIPLSKLFNDASLRAHRRLQLAFDTYQFFEAYIPKEUKYSFLONPQTSLSFSESIFT 86  
QY (?) PSNREETOQKSNLELLRISLLLSQSWLEPVOLGIGRFFVNHQUGS----- -HLVE 111  
DDE 87 PSNREETOQKSNLELLRISLLLSQSWLEPVOLGIGRFFVNHQUGS----- -HLVE 145  
QY 112 ALYLCVG--ERGFYTPKTRILVQ 134  
DB 146 GIOTLMGMLEDC---SPRTQIQPKQ 167

```

? CURRENT FILING DATE: 2002-12-13
? PRIOR APPLICATION NUMBER: US 60/248,478
? PRIOR FILING DATE: 2000-11-14
? PRIOR APPLICATION NUMBER: US 60/237,929
? PRIOR FILING DATE: 2000-10-02
? NUMBER OF SEQ ID NOS: 115
? SOFTWARE: PatentIn version 3.0
? SEQ ID NO 4
? LENGTH: 217
? TYPE: PRT
? ORGANISM: Homo sapiens
US-09-969-748C-4

```

	Query Match	Best Local Similarity	Score	DB 12:	Length	DB 16:	Gaps
Matches	101;	Conservative	7;	Mismatches	21;	Indels	16;
QY	2	FPTPLSLFDNMLRAHRLHQLAFD	YQFEFAY	IPKEQKYSFLQNPOTSLSFSES	1PT 61		
DB	27	FPTPLSLFDNMLRAHRLHQLAFD	YQFEFAY	IPKEQKYSFLQNPOTSLSFSES	1PT 86		
QY	62	PSNREETQOKSNLELLRISLLLI	QSWLDPVOLGTPRFVNHLCGS	-----	HLVE	111	
DB	87	PSNREETQOKSNLELLRISLLLI	QSWLDPVOLGTPRFVNHLCGS	-----	HLVE	145	

QY 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134  
DB 146 GIOTLMGRLEDG---SPRTGQIFKQ 167

RESULT 14  
US-09-804-409A-16  
; Sequence 16, Application US/09804409A  
; Patent No. US20020155100A1  
; GENERAL INFORMATION:  
; APPLICANT: KIEFFER, TIMOTHY J.  
; APPLICANT: CHEUNG, ANTHONY I.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATED PROTEIN  
; TITLE OF INVENTION: EXPRESSION IN GUT  
; FILE REFERENCE: 029996/027 8721  
; CURRENT APPLICATION NUMBER: US/09/804,409A  
; CURRENT FILING DATE: 2001-03-12  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 16  
; LENGTH: 217  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-804-409A-16

Query Match 57.0%; Score 454; DB 10; Length 217;  
Best Local Similarity 69.7%; Pred. No. 2.8e-42;  
Matches 101; Conservative 7; Mismatches 21; Indels 16; Gaps 4;  
QY 2 FPTIPLSRFDNAMLRAHLHQLAFTDYOEFEEAIPKEQKYSFLQNPOTSLSFSESIPT 61  
DB 27 FPTIPLSRFDNAMLRAHLHQLAFTDYOEFEEAIPKEQKYSFLQNPOTSLSFSESIPT 86  
QY 62 PSNREETOOKSNLELLRLISLLIQSWLEPVLGTGPRFVNQHLGCS-----HLVE 111  
DB 87 PSNREETOOKSNLELLRLISLLIQSWLEPVLGTGPRFVNQHLGCS-----HLVE 145  
QY 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134  
DB 146 GIOTLMGRLEDG---SPRTGQIFKQ 167

RESULT 15  
US-09-853-688-4  
; Sequence 4, Application US/09853688  
; Patent No. US20020081605A1  
; GENERAL INFORMATION:  
; APPLICANT: COOPER, DAVID N.  
; APPLICANT: PROCTER, ANNIE M.  
; APPLICANT: GREGORY, JOHN  
; APPLICANT: MILLAR, DAVID S.  
; TITLE OF INVENTION: METHOD FOR DETECTING GROWTH HORMONE VARIATIONS IN  
; FILE REFERENCE: WCH78  
; CURRENT APPLICATION NUMBER: US/09/853,688  
; CURRENT FILING DATE: 2001-05-14  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 217  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-853-688-4

Query Match 56.2%; Score 448; DB 9; Length 217;  
Best Local Similarity 69.0%; Pred. No. 1.3e-41;  
Matches 100; Conservative 7; Mismatches 22; Indels 16; Gaps 4;  
QY 2 FPTIPLSRFDNAMLRAHLHQLAFTDYOEFEEAIPKEQKYSFLQNPOTSLSFSESIPT 61  
DB 27 FPTIPLSRFDNAMLRAHLHQLAFTDYOEFEEAIPKEQKYSFLQNPOTSLSFSESIPT 86

Search completed: September 16, 2003, 12:52:26  
Job time : 47.2763 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 12:36:15 : Search time 17.5097 Seconds  
(without alignments)  
823.845 Million cell updates/sec

Title: US-09-423-100-7

Perfect score: 797

Sequence: 1 MFPTIPLSRLFDNMLRAHR.....IVEQCISICSLYLENYCN 150

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	461	57.8	217	1	STHU
2	460	57.7	217	1	somatotropin 1 pre
3	426.5	53.5	217	1	STHUV
4	426.5	53.5	256	1	somatotropin 2 pre
5	407.5	51.1	217	2	STHUV2
6	405	50.8	217	2	somatotropin 2 pre
7	396	49.7	212	2	chorionic somatoma
8	396	49.7	212	2	somatotropin - the
9	381	47.8	217	1	chorionic somatoma
10	381	47.8	217	1	chorionic somatoma
11	359.5	45.1	215	2	chorionamniotrophin
12	310.5	39.0	216	2	chorionamniotrophin
13	307.5	38.6	190	2	somatotropin - gol
14	306.5	38.5	190	1	somatotropin - sel
15	304.5	38.2	216	1	somatotropin - hor
16	302.5	38.0	216	1	somatotropin pre
17	302.5	38.0	216	1	somatotropin pre
18	301.5	37.8	190	2	somatotropin pre
19	301.5	37.8	216	1	somatotropin - Afr
20	301.5	37.8	216	1	STPG
21	301.5	37.8	216	2	somatotropin pre
22	299.5	37.6	216	2	somatotropin pre
23	297.5	37.3	190	2	somatotropin pre
24	295.5	37.1	190	2	somatotropin - alp
25	289.5	36.3	217	1	somatotropin - Arc
26	289.5	36.3	217	1	STBO
27	289.5	36.3	217	1	STSH
28	289.5	36.3	217	1	STGT
29	278.5	34.9	216	2	somatotropin - dom
					somatotropin pre

#### ALIGNMENTS

##### RESULT 1

STHU

somatotropin 1 precursor [validated] - human

N:Alternate names: growth hormone 1; hGH-N; pituitary somatotrophin

N:Contains: growth hormone 5K peptide; somatotropin 1, long form; somatotropin 1, shc

C:Species: Homo sapiens (man)

C:Date: 24-Apr-1984 #sequence revision 10-Feb-1995 #text change 08-Dec-2000

C:Accession: A93731; A32435; A53694; A94247; A90051; A93397; A93778; A91764; A90217;

R:DeNoto, F.M.; Moore, D.D.; Goodman, H.M.

Nucleic Acids Res. 9, 3719-3730, 1981

A:Title: Human growth hormone DNA sequence and mRNA structure: possible alternative s

A:Reference number: A93731; MUID:82014939; PMID:6269091

A:Accession: A93731

A:Molecule type: DNA

A:Residues: 1-217 <DEN>

A:Cross-references: GB:V00520

A>Note: the 20K short form somatoplin lacks residues 58-72 (32-45 in the active hor

R:Chen, E.Y.; Liao, Y.C.; Smith, T.; Barrera-Saldana, H.A.; Gelinias, R.E.; Seeburg,

Genomics 4, 479-497, 1989

A:Title: The human growth hormone

A:Reference number: A32435; MUID:1607277; PMID:2744760

A:Accession: A32435

A:Molecule type: DNA

A:Residues: 1-217 <CHE>

A:Cross-references: GB:V00571; NID:J184148; PDB:AAA52549.1; PDB:AAA54149

R:Roskar, W.; Kung'u, E.

Nucleic Acids Res. 7, 305-326, 1979

A:Title: Molecular cloning and nucleotide sequence of the human growth hormone struct

A:Reference number: A93694; MUID:80044477; PMID:386281

A:Accession: A93694

A:Molecule type: mRNA

A:Residues: 1-217 <ROS>

A:Cross-references: GB:V00519

A>Note: 35-Pro was also found

R:Marital, J.A.; Halliwell, R.A.; Baxter, J.D.; Goodman, H.M.

Science 205, 602-607, 1979

A:Title: Human growth hormone: complementary DNA cloning and expression in bacteria.

A:Reference number: A94247; MUID:79203293; PMID:377496

A:Accession: A94247

A:Molecule type: mRNA

A:Residues: 1-217 <MAR>

R:Li, C.H.; Dixon, J.S.; Liu, Y. K.

Arch. Biochem. Biophys. 133, 70-91, 1969

A:Title: Human pituitary growth hormone. XIX. The primary structure of the hormone.

A:Reference number: A90048; MUID:69289202; PMID:5810834

A:Contents: annotation

R:Li, C.H.; Dixon, J.S.

Arch. Biochem. Biophys. 146, 233-236, 1971

A:Title: Human pituitary growth hormone. XXXII. The primary structure of the hormone:

A:Reference number: A90051; MUID:72143935; PMID:5144027

A:Accession: A90051

A:Molecule type: protein

insulin precursor  
insulin precursor  
somatotropin precur  
epidermal growth f  
insulin - sperm wh  
insulin - finback  
insulin - elephant  
insulin precursor  
insulin precursor  
insulin - hamster  
insulin precursor  
insulin - Egyptian  
somatotropin - gre  
insulin precursor  
insulin precursor  
insulin precursor

30 277.5 34.8 110 1 INRB  
31 277.5 34.8 110 2 B42179  
32 275.5 34.6 216 2 A60509  
33 275 34.5 96 2 PC7082  
34 273.5 34.3 51 1 L4HP  
35 273.5 34.3 51 1 L4HP  
36 273.5 34.3 51 1 INEL  
37 273.5 34.3 110 2 JQ0178  
38 272 34.1 110 2 A42179  
39 271.5 34.1 51 1 INHY  
40 270 33.9 110 1 IPHU  
41 268.5 33.7 51 1 INMSSP  
42 268.5 33.7 191 2 A60425  
43 267.5 33.6 51 2 A59151  
44 266.5 33.4 105 1 IPBO  
45 265.5 33.3 110 2 I48166

A:Residues: 27-94:96-217 <LIC>  
R:Niall, H.D.  
Nature New Biol. 230, 90-91, 1971  
A:Title: Revised primary structure for human growth hormone.  
A:Reference number: A93397; MUID:71139765; PMID:5279046  
A:Accession: A93397  
A:Molecule type: protein  
A:Residues: 27-51 <NTA>  
R:Niall, H.D.; Hodan, M.L.; Sauer, R.; Rosenblum, I.Y.; Greenwood, F.C.  
Proc. Natl. Acad. Sci. U.S.A. 68, 866-869, 1971  
A:Title: Sequences of pituitary and placental lactogenic and growth hormones: evolution  
A:Reference number: A93778; MUID:71153968; PMID:5279528  
A:Accession: A93778  
A:Molecule type: protein  
A:Residues: 119-120:157-159 <NT2>  
R:Niall, H.D.  
In Prolactin and Carcinogenesis, Proc. Fourth Tenovus Workshop Prolactin, Griffiths, K.,  
A:Title: The chemistry of the human lactogenic hormones.  
A:Reference number: A94427  
A:Contents: annotation: somatotropin revision  
R:Bewley, T.A.; Dixon, J.S.; Li, C.H.  
Int. J. Pept. Protein Res. 4, 281-287, 1972  
A:Title: Sequence comparison of human pituitary growth hormone, human chorionic somatomedin C, and human growth hormone.  
A:Reference number: A91764; MUID:73092028; PMID:4675454  
A:Accession: A91764  
A:Molecule type: protein  
A:Residues: 27-217 <BEW>  
R:Lewis, U.J.; Bonewald, L.F.; Lewis, L.J.  
Biochem. Biophys. Res. Commun. 92, 511-516, 1980  
A:Title: The 20,000-dalton variant of human growth hormone: location of the amino acid  
A:Reference number: A90217; MUID:80130196; PMID:7356479  
A:Contents: somatotropin, 20K short variant  
A:Accession: A90217  
A:Molecule type: protein  
A:Residues: 46-57:73-80 <LEW>  
R:Chapman, G.E.; Rogers, K.M.; Brittain, T.; Bradshaw, R.A.; Bates, O.J.; Turner, C.; Ca  
J. Biol. Chem. 256, 2395-2401, 1981  
A:Title: The 20,000 molecular weight variant of human growth hormone. Preparation and so  
A:Reference number: A92311; MUID:8117361; PM: 7462247  
A:Contents: somatotropin, 20K short variant  
A:Accession: A92311  
A:Molecule type: protein  
A:Residues: 27-57:73-79 <CHA>  
R:Singh, R.N.P.; Seavey, B.K.; Lewis, L.J.; Lewis, U.J.  
J. Protein Chem. 2, 425-436, 1983  
A:Title: Human growth hormone peptide 1-43: isolation from pituitary glands.  
A:Reference number: A61466  
A:Accession: A61466  
A:Molecule type: protein  
A:Residues: 27-69 <SIN>  
R:Robson, V.M.J.; Rae, I.D.; NG, F.  
Biol. Chem. Hoppe-Seyler 371, 423-431, 1990  
A:Title: Identification of the aspartimide structure in a previously-reported peptide.  
A:Reference number: S09685; MUID:90334745; PMID:2378679  
A:Accession: S09685  
A:Molecule type: protein  
A:Residues: 27-34,'L',36-47 <ROB>  
R:de Vos, A.M.; Ultsch, M.; Kossiakoff, A.A.  
Science 255, 306-312, 1992  
A:Title: Human growth hormone and extracellular domain of its receptor: crystal structure  
A:Reference number: A41728; MUID:92196577; PMID:1549776  
A:Contents: annotation; X-ray crystallography, 2.8 angstroms  
A:Note: the structure of the complex with growth hormone receptor is described  
R:Gray, G.L.; Balridge, J.S.; McKown, K.S.; Heyneker, H.L.; Chang, C.N.  
Gene 39, 247-254, 1985  
A:Title: Periplasmic production of correctly processed human growth hormone in Escherich  
A:Reference number: I41126; MUID:86137393; PMID:3912261  
A:Accession: I41549  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-26 <RES>  
A:CROSS-references: GB:M14398; NID:g183159; PIDN:AAA52554.1; PID:g183159

C:Comment: The gene for this hormone is transcribed only in somatotrophic cells of th  
C:Comment: About 90% of somatotropin is the 22K long form.  
C:Genetics:  
A:Gene: GDB:GH1  
A:CROSS-references: GDB:119982; OMIM:139250  
A:Map position: 17q23.1-17q23.3  
A:Introns: 4/1; 57/3; 97/3; 152/3  
C:Superfamily: prolactin  
C:Keywords: alternative splicing; hormone; pituitary  
F:1-26/Domain: signal sequence #status experimental <SIG>  
F:27-217/Product: somatotropin long form #status experimental <SOL>  
F:27-69/Product: growth hormone 5K peptide #status experimental <SKP>  
F:27-57,73-217/Product: somatotropin short form #status experimental <SOS>  
F:79-191,208-215/Disulfide bonds: #status experimental

Query Match 57.8% Score 461; DB 1; Length 217;  
Best Local Similarity 70.3%; Ed. No. 7.5e-38;  
Matches 102; Conservative 1; Mismatches 20; Indels 16; Gaps 4;  
QY 2 FPTIPLSRLFDNMLRAHLHLQHLAFDIYOFEEAYIPKEQKYSFLQNPTSLSFSESPT 61  
DB 27 FPTIPLSRLFDNMLRAHLHLQHLAFDIYOFEEAYIPKEQKYSFLQNPTSLSFSESPT 86  
QY 62 PSNREETQOKSNLELLRISLLLIQSNLEPVLQGTGPRFVNQHLGGS-----HLVE 111  
DB 87 PSNREETQOKSNLELLRISLLLIQSNLEPVLQGTGPRFVNQHLGGS-----HLVE 145

QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134  
DB 146 GIOTLMGRLEDG---SPRTGQIFKQ 167

RESULT 2  
167410  
somatotropin - rhesus macaque  
N:Alternate names: growth hormone  
C:Species: Macaca mulatta (rhesus macaque)  
C:Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 16-Jul-1999  
C:Accession: 167410; A05094  
R:Golos, T.G.; Darning, M.; Fisher, J.M.; Fowler, P.D.  
Endocrinology 135, 1744-1752, 1995  
A:Title: Cloning of four growth hormone/chorionic somatomamotropin-related complemen  
A:Reference number: 153267; MUID:64708724; PMID:8404617  
A:Accession: 167410  
A:Status: translated from GR/EMBL; 187  
A:Molecule type: mRNA  
A:Residues: 1-217 <RES>  
A:CROSS-references: GB:116566; NID:1243114; PDB:AAA16842.1; PDB:1243115  
R:Li, C.H.; Chung, D.; Lamm, H.W.; Stenke, S.  
Arch. Biochem. Biophys. 245, 287-291, 1986  
A:Title: The primary structure of monkey pituitary growth hormone.  
A:Reference number: A05094; MUID:86129460; PMID:3080959  
A:Accession: A05094  
A:Molecule type: protein  
A:Residues: 27-99,'Q',101-178,'D',180-217 <LIC>  
A:Note: the monkey species is not identified in the reference  
R:Raben, M.S.  
Science 125, 883-884, 1957  
A:Title: Preparation of growth hormone from pituitaries of man and monkey.  
A:Reference number: A44774  
A:Contents: annotation; identification of source organism  
C:Superfamily: prolactin

Query Match 57.7% Score 460; DB 2; Length 217;  
Best Local Similarity 98.9%; Ed. No. 9.4e-38;  
Matches 90; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 2 FPTIPLSRLFDNMLRAHLHLQHLAFDIYOFEEAYIPKEQKYSFLQNPTSLSFSESPT 61  
DB 27 FPTIPLSRLFDNMLRAHLHLQHLAFDIYOFEEAYIPKEQKYSFLQNPTSLSFSESPT 86  
QY 62 PSNREETQOKSNLELLRISLLLIQSNLEPVLQGTGPRFVNQHLGGS-----HLVE 92



A:Title: Cloning of four growth hormone/chorionic somatomotropin-related complementary DNAs from human placenta  
A:Accession: F53267  
A:Reference number: 153267; MUID:94006724; PMID:8404617  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-217 <RES>  
A:Cross-references: GB:L6552; NID:q293108; PION:AAA18839.1; PID:q293109  
C:Superfamily: prolactin

Query Match 49.7%; Score 396; DB 2; Length 217;  
Best Local Similarity 82.2%; Pred. No. 1.8e-31;  
Matches 74; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

QY 3 PTIPLSLFDNALRAHRLHQLAFDYQEEYAIPEKQKYSFLQNHQTSLSFSESIPTP 62  
|||||  
DB 28 PSVPLSLFDNALRAHRLHQLAFDYQEEYAIPEKKHSLMENPQASFCADSIPTP 87  
|||||  
QY 63 SNREETQOKSNLELLRTISLLLSQSWLEPVQ 92  
|||||  
DB 88 SNLEETQOKSNLELLRTISLLLSQSWLEPVQ 117  
|||||

RESULT 9  
LCHUC  
Choriomamotropin A precursor [validated] - human  
N:Alternate names: chorionic somatomamotropin 1; placental lactogen  
C:Species: Homo sapiens (man)  
C:Date: 23-Oct-1981; sequence revision 23-Oct-1981; text change 06-Dec-2000  
C:Accession: C32435; A94422; I52342; A91833; A93192; A90054; A94427; A61283; I55229;  
C:Chen, E.T.; Liao, Y.C.; Smith, D.H.; Barrera-Saidana, H.A.; Gelinas, R.E.; Seeburg,  
Genomics 4, 479-497, 1989  
A:Title: The human growth hormone locus: nucleotide sequence, biology, and evolution  
A:Reference number: A32435; MUID:89307277; PMID:2744760  
A:Accession: C32435  
A:Molecule type: DNA  
A:Residues: 1-217 <RES>  
A:Cross-references: GB:J03071; NID:J183148; PION:AAAS2551.1; PID:q181151  
R:Goodman, H.M.; Denoto, F.; Fiddes, J.C.; Halliwell, R.A.; Page, G.S.; Smith, S.; Tili  
in Mobilization and Reassembly of Genetic Information, Scott, W.A.; Weiner, R.; Joseph  
A:Reference number: A94422  
A:Accession: A94422  
A:Molecule type: mRNA  
A:Residues: 1-217 <GOO>  
R:Tanaka, M.; Masuda, N.; Watabiki, M.; Yamakawa, M.; Shimizu, K.; Nasai, J.; Nakashi  
Biochem. Int. 16, 287-292, 1988  
A:Title: cDNA cloning of human chorionic somatomamotropin-1 mRNA whose transcription  
A:Reference number: I52342; MUID:88279366; PMID:2815350  
A:Accession: I52342  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-3 <IANS>  
A:Cross-references: GB:M35419; NID:q506822  
R:Sherwood, L.M.; Burstein, Y.; Schechter, I.  
Proc. Natl. Acad. Sci. U.S.A. 76, 3819-3823, 1979  
A:Title: Primary structure of the NH-2-terminal extra piece of the precursor to human  
A:Reference number: A93833; MUID:90034970; PMID:291043  
A:Accession: A93833  
A:Molecule type: protein  
A:Residues: 1-3-26 <SHE>  
A:Experiment 1 source: placenta  
R:Shine, J.; Seeburg, P.H.; Martla, J.A.; Baxter, J.D.; Goodman, H.M.  
Nature 270, 494-499, 1977  
A:Title: Construction and analysis of recombinant DNA for human chorionic somatomam  
A:Reference number: A93192; MUID:7801761; PMID:593368  
A:Accession: A93192  
A:Molecule type: DNA  
A:Residues: 50-217 <SHI>  
R:Li, C.H.; Dixon, J.S.; Chung, D.  
Arch. Biochem. Biophys. 155, 95-110, 1973  
A:Title: Amino acid sequence of human chorionic somatomamotropin.  
A:Reference number: A90054; MUID:73201971; PMID:4712450  
A:Accession: A90054



```

QY 64 NREETQOKSNNLELLRISLLLIQSWLEPVQ 92
DB 89 NMEEETQOKSNNLELLRISLLLIQSWLEPVR 117

RESULT 10
E32435
Chorionamniotropin B precursor - human
N:Alternate names: chorionic somatomammotropin 2
C:Species: Homo sapiens (man)
C:Date: 29-Dec-1989 #sequence_revision 29-Dec-1989 #text_change 16-Jul-1989
C:Accession: 32435
R:Chen, E.Y.; Liao, Y.C.; Smith, D.H.; Bartera-Saldana, H.A.; Gelinas, R.E.; Seeburg,
Genomics 4, 479-497, 1989
A:Title: The human growth hormone locus: nucleotide sequence, biology, and evolution
A:Reference number: A32435; MUID:89307277; PMID:2744760
A:Accession: E32435
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-217 <CHE>
A:Cross-references: GB:J03071; NID:q183148; PIDN:AAA52553.1; PID:q183153
C:Genetics:
A:Gene: GDB:CSH2
A:Cross-references: GDB:119813; OMIM:118820
A:Map position: 17q22-17q24
C:Superfamily: prolactin

Query Match 47.8%; Score 381; DB 2; Length 217.
Best local Similarity 82.0%; Pred. No. 5.4e-30;
Matches 73; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 4 TPLSLRFNAMLRARHLHQLAFDTYQEEFAEYIPKEQKYSFLQNPOTLSLSFSESIPTPS 63
DB 29 TVPLSLRFNAMLRARHLHQLAFDTYQEEFAEYIPKEQKYSFLHDSQTSFCFSDSIPTPS 88

QY 64 NREETQOKSNNLELLRISLLLIQSWLEPVQ 92
DB 89 NMEEETQOKSNNLELLRISLLLIQSWLEPVR 117

RESULT 11
E32649
Chorionamniotropin precursor (allele hCS-3) - human
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 28-Jul-1995
C:Accession: A26449
R:Hiatt, H.; Kimmelman, J.; Birnbaumer, M.J.; Chen, E.Y.; Seeburg, P.H.; Eberhardt, P.L.;
DNA 6, 59-70, 1987
A:Title: The human growth hormone gene locus: structure, evolution, and allelic varia
A:Reference number: A26449; MUID:87161215; PMID:3030680
A:Accession: A26449
A:Molecule type: DNA
A:Residues: 1-215 <HIR>
C:Superfamily: prolactin
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-215/Product: chorionamniotropin, hCS-3 allele #status predicted <MAT>

Query Match 45.1%; Score 359.5; DB 2; Length 215;
Best local Similarity 80.5%; Pred. No. 7e-28;
Matches 70; Conservative 8; Mismatches 8; Indels 1; Gaps 1;

QY 4 TPLSLRFNAMLRARHLHQLAFDTYQEEFAEYIPKEQKYSFLQNPOTLSLSFSESIPTPS 63
DB 29 TVPLSLRFNAMLRARHLHQLAFDTYQEEFAEYIPKEQKYSFLHDSQTSFCFSDSIPTPS 88

QY 64 NREETQOKSNNLELLRISLLLIQSWLEP 90
DB 89 NMEEETQOKSNNLELLRISLLLIQSWLEP 114

RESULT 12
B49159
somatotropin - golden hamster

```

```

A:Molecule type: protein
A:Residues: 1-190 <ZAK>
R:Zakin, M.M.; Poskus, E.; Dellachia, J.M.; Paladini, A.C.; Santome, J.A.
FEBS Lett. 34, 353-355, 1973
A:Title: The amino acid sequence of equine growth hormone.
A:Reference number: A91395; MUID:71020362; PMID:4747849
A:Accession: A91395
A:Molecule type: protein
A:Residues: 1-190 <ZA>
R:Zakin, M.M.; Poskus, E.; Dellachia, J.M.; Paladini, A.C.; Santome, J.A.
FEBS Lett. 25, 77-82, 1972
A:Title: Amino acid sequences around the cysteine residues in equine growth hormone.
A:Reference number: A91383
A:Accession: A91383
A:Molecule type: protein
A:Residues: 42-69,157-190 <ZA>
R:Oliver, L.; Hartree, A.S.
Biochem. J. 109, 19-24, 1968
A:Title: Amino acid sequences around the cysteine residues in horse growth hormone.
A:Reference number: A90240; MUIH:48168190; PMID:4876100
A:Accession: A90240
A:Molecule type: protein
A:Residues: 176-190 <OLI>
C:Superfamily: prolactin
C:Keywords: hormone; pituitary
F:52-163,180-188/Disulfide bonds: *status experimental

Query Match      38.5%   Score 306.5; DB 1; Length 190;
Best Local Similarity 55.2%   ; ed. No. 9,7e-23;
Matches 60; Conservative 14 Mismatches 17; Indels 1; Gaps 1;

QY    2 FPTPLSLFDNAMLRAHRHLHOLAFDTYQFFEEAVIPKPKYSFLQNPNQISLFSSESIFT 61
DB    1 FPANPLSSLFANVLRAQHILHOLAADTYKEFERATPEQQRYS-IQNQAACFCSETIPA 59

QY    62 PSNREETQOKSNLELLRLSILLIQSWSLPVQL 93
DB    60 PTGDEAQQRSDMELLRFSLILLSWLGVPVQL 91

RESULT 15
SUMS
somatotropin precursor : mouse
N:Alternate names: growth hormone
C:Species: Mus musculus (house mouse)
C>Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #ext_change 28-May-1999
C:Accession: R24911
R:Linzer, D.I.H.; Talamantes, F.
J. Biol. Chem. 260, 5574-5579, 1985
A:Title: Nucleotide sequence of mouse prolactin and growth hormone mRNAs and expression of cDNA clones encoding them.
A:Reference number: A92548; MUIB:8526358; PMID:2991252
A:Accession: R23911
A:Molecule type: mRNA
A:Residues: 1-216 <LIN>
A:Cross-references: GB:X02891; GB:X03232; MID:951067; PIDN:CAA26650.1; PID:g51068
C:Superfamily: prolactin
C:Keywords: anterior pituitary; growth factor; hormone
F:1-26/DNA; signal sequence
F:27-216/Product; somatotropin
F:78-189,206-214/Disulfide bonds: *status predicted

Query Match      38.2%   Score 304.5; DB 1; Length 216;
Best Local Similarity 54.8%   ; Pred. No. 1,8e-22;
Matches 59; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

QY    2 FPTPLSLFDNAMLRAHRHLHOLAFDTYQFFEEAVIPKPKYSFLQNPNQISLFSSESIFT 61
DB    27 FPANPLSSLFANVLRAQHILHOLAADTYKEFERATPEQQRYS-IQNQAACFCSETIPA 85

QY    62 PSNREETQOKSNLELLRLSILLIQSWSLPVQ 92
DB    86 PTGDEAQQRSDMELLRFSLILLSWLGVPVQ 116

```

Search completed: September 16, 2003, 12:40:36  
Job time : 18.5097 secs

---

GenCore version 1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 12:33:30 : Search time 11.6732 Seconds  
(without alignments)  
604.293 Million cell updates/sec

Title: us-09-423-100-7

Perfect score: 797

Sequence: 1 MFPTIPLSRLEFDNMLRAHR.....IVEQCCTICSLSYQLENVCN 150

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	461	57.8	217	1 SOMA_HUMAN	P01241 homo sapien
2	461	57.8	217	1 SOMA_PANTR	P58756 pan troglod
3	460	57.7	217	1 SOMA_MACMU	P33093 macaca mula
4	437	54.8	217	1 SOMA_SAIIB	P58343 salmirl bol
5	434.5	54.5	217	1 SOM2_PANTR	P58757 pan troglod
6	432	54.2	217	1 SOMA_CALJA	Q9amb3 callithrix
7	426.5	53.5	217	1 SOM2_HUMAN	P01242 homo sapien
8	399	50.1	217	1 SOM2_MACMU	Q07370 macaca mula
9	381	47.8	217	1 PLL_HUMAN	P01243 homo sapien
10	310.5	39.0	216	1 SOMA_MESAU	P37886 mesocricetu
11	307.5	38.6	190	1 SOMA_BALBO	P33092 balaeopter
12	306.5	38.5	216	1 SOMA_HORSE	P01245 equus cabal
13	306.5	38.5	217	1 SOMA_GALSE	Q9qkal galago sene
14	306.5	38.5	217	1 SOMA_NYCPY	Q9amb2 nycticebus
15	304.5	38.2	216	1 SOMA_MOUSE	P06880 mus musculu
16	302.5	38.0	216	1 SOMA_RABIT	P46407 oryctolagus
17	302.5	38.0	216	1 SOMA_RAT	P01244 rattus norv
18	301.5	37.8	190	1 SOMA_LOXAF	P20392 loxodonta a
19	301.5	37.8	216	1 SOMA_CANFA	P33711 canis fami
20	301.5	37.8	216	1 SOMA_FELCA	P46404 felis silve
21	301.5	37.8	216	1 SOMA_PIG	P01248 sus scrofa
22	299.5	37.6	216	1 SOMA_MUSVI	P19795 mustela vis
23	297.5	37.3	190	1 SOMA_LAMPA	P37885 lama guanac
24	295.5	37.1	190	1 SOMA_VULVU	P10766 vulpes vulp
25	291.5	36.6	215	1 SOMA_MONDO	Q9g160 monodelphis
26	291.5	36.6	215	1 SOMA_TRIVU	O62754 trichosurus
27	289.5	36.3	217	1 SOMA_BOVIN	P01246 bos taurus
28	289.5	36.3	217	1 SOMA_CEREL	P56437 cervus elap
29	289.5	36.3	217	1 SOMA_SHEEP	P01247 ovis aries
30	282.5	35.4	217	1 SOMA_BUBBU	O18938 bubalus bub
31	278.5	34.9	216	1 SOMA_MELGA	P22077 meleagris g
32	277.5	34.8	110	1 INS_CERAE	P30407 cercopithec
33	277.5	34.8	110	1 INS_RABIT	P01311 oryctolagus

#### RESULT 1

ID	SOMA_HUMAN	STANDARD	PRT	217 AA
AC	P01241: Q14405; Q16631; Q9HB21; Q9UMJ7; Q9UNL5;			
DI	21-JUL-1986 (Rel. 01, Created)			
DT	01-MAR-1992 (Rel. 21, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Somatotropin precursor (Growth hormone) (GH) (GH-N) (Pituitary growth hormone) (Growth hormone 1).			
GN	GH1			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RX	MEDLINE=80034477; PubMed=386281;			
RA	Roskam W., Rougeon F.;			
RT	"Molecular cloning and nucleotide sequence of the human growth hormone structural gene."			
RT	hormone			
RL	Nucleic Acids Res. 7:305-320(1979).			
RN	[2]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RX	MEDLINE=79203293; PubMed=377496;			
RA	Martial J.A., Halliwell R.A., Baxter J.D., Goodman H.M.;			
RT	"Human growth hormone: complementary DNA cloning and expression in bacteria."			
RL	Science 205:602-607(1979).			
RN	[3]			
RP	SEQUENCE FROM N.A. (ISOFORM 1), AND POSSIBLE ALTERNATIVE SPLICING.			
RX	MEDLINE=82014339; PubMed=6264691;			
RA	Denoto F.M., Moore D.D., Goodman H.M.;			
RT	"Human growth hormone DNA sequence and mRNA structure: possible alternative splicing."			
RL	Nucleic Acids Res. 9:3719-3730(1981).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=83182010; PubMed=7169009;			
RA	Seeburg P.H.;			
RT	"The human growth hormone gene family: nucleotide sequences show recent divergence and predict a new polypeptide hormone."			
RL	DNA 1:239-249(1982).			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=89307277; PubMed=274 760;			
RA	Chen E.Y., Liao Y.C., Smith J.H., Barrera-Saidana H.A.,			
RA	Gelinas R.E., Seeburg P.H.;			
RT	"The human growth hormone locus: nucleotide sequence, biology, and evolution."			
RL	Genomics 4:479-497(1989).			
RN	[6]			
RP	SEQUENCE FROM N.A. (ISOFORM 3).			
RC	TISSUE=Pituitary;			
RA	Gu J., Huang Q.-H., Li N., Xu S.-H., Han Z.-G., Fu G., Chen Z.;			
RT	"A novel gene expressed in human pituitary."			
RT	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.			

P08998 gallus galli  
Q9hwq3 struthio ca  
P01312 balaeopter  
P01316 eiephas max  
P30406 macaca fasc  
P55755 crocodylus  
P30410 pan troglod  
P01308 homo sapien  
P01324 acromys capl  
P34005 cheloniana my  
P01317 bos taurus  
Q9ixi3 spermophilu

#### ALIGNMENTS

- RN [7] SEQUENCE FROM N.A. (ISOFORM 4).  
 RP TISSUE-Pituitary; PubMed-10931946;  
 RC MEDLINE-20402571;  
 RX Hu R.-M., Han Z.-G., Song H.-D., Peng Y.-D., Huang Q.-H., Ren S.-X.,  
 RA Gu Y.-J., Huang C.-H., Li Y.-B., Jiang C.-L., Fu G., Zhang Q.-H.,  
 RA Gu B.-W., Dai M., Mao Y.-F., Gao G.-F., Rong R., Ye M., Zhou J.,  
 RA Xu S.-H., Gu J.-X., Shi J.-X., Jin W.-R., Zhang C.-K., Wu T.-M.,  
 RA Huang G.-Y., Chen Z., Chen M.-D., Chen J.-L.;  
 RT \*Gene expression profiling in the human hypothalamus-pituitary-adrenal  
 RT axis and full-length cDNA cloning.\*;  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:9543-9548(2000).  
 RN [18]  
 RP SEQUENCE OF 1-26 FROM N.A.  
 RX MEDLINE-86137393; PubMed-3912261;  
 RA Gray G.L., Baldrige J.S., McKeown K.S., Heyneker H.L., Chang C.N.;  
 RT \*Periplasmic production of correctly processed human growth hormone in  
 RT Escherichia coli; natural and bacterial signal sequences are  
 RT interchangeable.\*;  
 RL Gene 39:247-254(1985).  
 RN [19]  
 RP SEQUENCE OF 27-217.  
 RX MEDLINE-69289202; PubMed-5810834;  
 RA Li C.H., Dixon J.S., Liu W.-K.;  
 RT \*Human pituitary growth hormone. XIX. The primary structure of the  
 RT hormone.\*;  
 RL Arch. Biochem. Biophys. 133:70-91(1969).  
 RN [10]  
 RP SEQUENCE OF 27-217, AND REVISIONS.  
 RX MEDLINE-72143935; PubMed-5144027;  
 RA Li C.H., Dixon J.S.;  
 RT \*Human pituitary growth hormone. 32. The primary structure of the  
 RT hormone; revision.\*;  
 RL Arch. Biochem. Biophys. 146:233-236(1971).  
 RN [11]  
 RP REVISION.  
 RX MEDLINE-73092028; PubMed-4675454;  
 RA Bewley T.A., Dixon J.S., Li C.H.;  
 RT \*Sequence comparison of human pituitary growth hormone, human  
 RT chorionic somatomotropin, and ovine pituitary growth and  
 RT lactogenic hormones.\*;  
 RL Int. J. Pept. Protein Res. 4:281-287(1972).  
 RN [12]  
 RP SEQUENCE OF 27-61 AND 102-124.  
 RX MEDLINE-71139765; PubMed-5279046;  
 RA Niall H.D.;  
 RT \*Revised primary structure for human growth hormone.\*;  
 RL Nature New Biol. 230:90-91(1971).  
 RN [13]  
 RP REVISIONS TO 119-120 AND 157-159.  
 RX MEDLINE-71153968; PubMed-5279528;  
 RA Niall H.D., Hogan M.L., Sauer R., Rosenbium I.Y., Greenwood F.C.;  
 RT \*Sequences of pituitary and placental lactogenic and growth hormones:  
 RT evolution from a primordial peptide by gene reduplication.\*;  
 RL Proc. Natl. Acad. Sci. U.S.A. 68:866-869(1971).  
 RN [14]  
 RP REVISION.  
 RA Niall H.D.;  
 RT \*The chemistry of the human lactogenic hormones.\*;  
 RL Prolactin and carcinogenesis, proc. fourth tenovus workshop prolactin,  
 RL pp.13-20, Alpha Omega Alpha Press, Cardiff (1972).  
 RN [15]  
 RP SEQUENCE OF 27-79 (ISOFORM 2).  
 RX MEDLINE-81117361; PubMed-7462247;  
 RA Chapman G.E., Rogers K.M., Brittain T., Bradshaw R.A., Bates O.J.,  
 RA Turner C., Cary P.D., Crane-Robinson C.;  
 RT \*The 20,000 molecular weight variant of human growth hormone.  
 RT Preparation and some physical and chemical properties.\*;  
 RL J. Biol. Chem. 256:2395-2401(1981).  
 RN [16]  
 RP SEQUENCE OF 46-80 (ISOFORM 2).  
 RX MEDLINE-80130196; PubMed-7356479;  
 RA Lewis U.J., Bonewald L.F., Lewis L.J.;  
 RT \*The 20,000-dalton variant of human growth hormone: location of the  
 RT amino acid deletions.\*;  
 RL Biochem. Biophys. Res. Commun. 92:511-516(1980).  
 RN [17]  
 RP DEAMIDATION OF GLN-163 AND ASN-178.  
 RX MEDLINE-82052997; PubMed-7028740;  
 RA Lewis U.J., Singh R.N., Bonewald L.F., Seavey B.K.;  
 RT \*Altered proteolytic cleavage of human growth hormone as a result of  
 RT deamidation.\*;  
 RL J. Biol. Chem. 256:11645-11650(1981).  
 RN [18]  
 RP REVIEW.  
 RX MEDLINE-99321812; PubMed-10193484;  
 RA Baumann G.;  
 RT \*Growth hormone heterogeneity in human pituitary and plasma.\*;  
 RL Horm. Res. 51 Suppl. 1:2-6(1999).  
 RN [19]  
 RP 3D-STRUCTURE MODELING.  
 RX MEDLINE-88190073; PubMed-3447173;  
 RA Cohen F.E., Kuntz I.D.;  
 RT \*Prediction of the three-dimensional structure of human growth  
 RT hormone.\*;  
 RL Proteins 2:162-166(1987).  
 RN [20]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
 RX MEDLINE-92196577; PubMed-1549776;  
 RA de Vos A.M., Uitsch M., Kossiakoff A.A.;  
 RT \*Human growth hormone and extracellular domain of its receptor:  
 RT crystal structure of the complex.\*;  
 RL Science 255:306-312(1992).  
 RN [21]  
 RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).  
 RX MEDLINE-95075452; PubMed-7984244;  
 RA Somers W., Uitsch M., de Vos A.M., Kossiakoff A.A.;  
 RT \*The X-ray structure of a growth hormone-prolactin receptor complex.\*;  
 RL Nature 371:478-481(1994).  
 RN [22]  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
 RA Chantalat L., Chirgadze N.Y., Jones N., Korber F., Navaza J.,  
 RA Pavlovsk A.G., Wlodawer A.;  
 RT \*The crystal-structure of wild-type growth-hormone at 2.5-A  
 RT resolution.\*;  
 RL Protein Pept. Lett. 2:333-340(1995).  
 RN [23]  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
 RX MEDLINE-97113024; PubMed-4943273;  
 RA Sundstroem M., Lundqvist T., Rodin J., Giebel L.B., Milligan D.,  
 RA Norstedt G.;  
 RT \*Crystal structure of an antagonist mutant of human growth hormone,  
 RT G120R, in complex with its receptor at 2.9-A resolution.\*;  
 RL J. Biol. Chem. 271:32197-32203(1996).  
 RN [24]  
 RP VARIANT CYS-105.  
 RX MEDLINE-99318093; PubMed-10391209;  
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nemesh J., Ziaugra L.,  
 RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,  
 RA Lander E.S.;  
 RT \*Characterization of single-nucleotide polymorphisms in coding regions  
 RT of human genes.\*;  
 RL Nat. Genet. 22:231-238(1999).  
 RN [25]  
 RP ERRATUM.  
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nemesh J., Ziaugra L.,  
 RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,  
 RA Lander E.S.;  
 RL Nat. Genet. 23:373-373(1999).  
 CC -1- FUNCTION: Plays an important role in growth control. Its major  
 CC role in stimulating body growth is to stimulate the liver and  
 CC other tissues to secrete IGF-1. It stimulates both the  
 CC differentiation and proliferation of myoblasts. It also stimulates

Query Match 57.8%; Score 461; DB 1; Length 217;  
 Best Local Similarity 70.3%; Pred. No. 4.6e-38;  
 Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;

Qy 2 FPIPLSRFDNMLRAHRLHOLAFDTYQFEAYIPKEQKYSFLONPOTSLSFSESIPT 61  
 |||||  
 Db 27 FPIPLSRFDNMLRAHRLHOLAFDTYQFEAYIPKEQKYSFLONPOTSLSFSESIPT 86  
 |||||  
 Qy 62 PSNRETOQKSNLELLRISLLLIQSWLEPVQGTGPRFVNOHLGCS-----HLVE 111  
 |||||  
 Db 87 PSNRETOQKSNLELLRISLLLIQSWLEPVQGTGPRFVNOHLGCS-----HLVE 145  
 |||||  
 Qy 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134  
 : : : : :  
 Db 146 GIOTLMGRLEDG---SPRTGQIFKQ 167  
 : : : : :  
 |||||

## RESULT 2

SOMA\_PANTR ID SOMA\_PANTR STANDARD; PRT: 217 AA.  
 AC P58756;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Somatotropin precursor (Growth hormone) (GH) (GH-N) (Pituitary growth hormone) (Growth hormone 1).  
 GN GH1.  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 OX NCBI\_TaxID=9598;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Revol A., Esquivel D., Santiago D., Barrera-Saldana H.;  
 RT "Independent duplication of the growth hormone gene in three Anthropoid lineages."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Plays an important role in growth control. Its major role in stimulating body growth is to stimulate the liver and other tissues to secrete IGF-1. It stimulates both the differentiation and proliferation of myoblasts. It also stimulates amino acid uptake and protein synthesis in muscle and other tissues (BY similarity).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: AF374232; AAL72284.1;  
 CC InterPro: IPR001400; Somatotropin.  
 CC Pfam: PF00103; hormone; 1.  
 CC PRINTS: PR00836; SOMATOTROPIN.  
 CC PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 CC PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 CC Hormone; Pituitary; Signal.  
 CC SIGNAL 1 26 BY SIMILARITY.  
 CC CHAIN 27 217 SOMATOTROPIN.  
 CC DISULFID 79 191 BY SIMILARITY.  
 CC DISULFID 208 215 BY SIMILARITY.  
 CC SEQUENCE 217 AA; 24843 MW; FEA295DE0518674 CRC64;  
 Query Match 57.8%; Score 461; DB 1; Length 217;  
 Best Local Similarity 70.3%; Pred. No. 4.6e-38;  
 Matches 102; Conservative 7; Mismatches 20; Indels 16; Gaps 4;

Qy 2 FPIPLSRFDNMLRAHRLHOLAFDTYQFEAYIPKEQKYSFLONPOTSLSFSESIPT 61

Db 27 FPIPLSRFDNMLRAHRLHOLAFDTYQFEAYIPKEQKYSFLONPOTSLSFSESIPT 86  
 |||||  
 Qy 62 PSNRETOQKSNLELLRISLLLIQSWLEPVQGTGPRFVNOHLGCS-----HLVE 111  
 |||||  
 Db 87 PSNRETOQKSNLELLRISLLLIQSWLEPVQGTGPRFVNOHLGCS-----HLVE 145  
 |||||  
 Qy 112 ALYLVCG--ERGFFYTPKTRGIVEQ 134  
 : : : : :  
 Db 146 GIOTLMGRLEDG---SPRTGQIFKQ 167  
 : : : : :  
 |||||

## RESULT 3

SOMA\_MACMU ID SOMA\_MACMU STANDARD; PRT: 217 AA.  
 AC P33093;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Somatotropin precursor (Growth hormone) (GH) (GH-N) (Pituitary growth hormone) (Growth hormone 1).  
 GN GH1.  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Macaca.  
 OX NCBI\_TaxID=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=94008724; PubMed=8404.17;  
 RT "Cloning of four growth hormone/chorionic somatomammotropin-related complementary deoxyribonucleic acids differentially expressed during pregnancy in the rhesus monkey placenta."  
 RL Endocrinology 133:1744-1752 (1993).  
 RN [2]  
 RP SEQUENCE OF 27-217.  
 RA MEDLINE=86129460; PubMed=3080959;  
 RA Li C.H., Chung D., Lahn H.W., Stein S.;  
 RT "The primary structure of monkey pituitary growth hormone."  
 RL Arch. Biochem. Biophys. 245:287-291 (1986).  
 CC -1- FUNCTION: Plays an important role in growth control. Its major role in stimulating body growth is to stimulate the liver and other tissues to secrete IGF-1. It stimulates both the differentiation and proliferation of myoblasts. It also stimulates amino acid uptake and protein synthesis in muscle and other tissues.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: L16556; AAA18842.1;  
 CC PIR: I67410; I67410.  
 CC HSSP: P01241; IAXI.  
 CC InterPro: IPR001400; Somatotropin.  
 CC Pfam: PF00103; hormone; 1.  
 CC PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 CC PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 CC Hormone; Pituitary; Signal.  
 CC SIGNAL 1 26 BY SIMILARITY.  
 CC CHAIN 27 217 SOMATOTROPIN.  
 CC DISULFID 79 191 BY SIMILARITY.  
 CC DISULFID 208 215 BY SIMILARITY.  
 CC CONFLICT 100 100 E -> Q (IN REF. 2).  
 CC CONFLICT 179 179 N -> D (IN REF. 2).  
 CC SEQUENCE 217 AA; 24913 MW; 2C5180341EEC46D0 CRC64;

Query Match 57.7%; Score 460; DB 1; Length 217;  
 Best Local Similarity 98.9%; Pred. No. 5.8e-38;  
 Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FPTPLSLFONAMLRHRLHQLAFDTYQEEAYIPKEQKYSFLONPQTSLSFSESPT 61  
 ID 27 FPTPLSLFONAMLRHRLHQLAFDTYQEEAYIPKEQKYSFLONPQTSLSFSESPT 86  
 DB 27 FPTPLSLFONAMLRHRLHQLAFDTYQEEAYIPKEQKYSFLONPQTSLSFSESPT 86

QY 62 PSNRETOOKSNLELLRISLLLIQSWLEPVQ 92  
 ID 87 PSNRETOOKSNLELLRISLLLIQSWLEPVQ 117  
 DB 87 PSNRETOOKSNLELLRISLLLIQSWLEPVQ 117

RESULT 4  
 SOMA\_SAIBB STANDARD: PRT: 217 AA.  
 AC P58343;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Somatotropin precursor (Growth hormone).  
 GN CH1.  
 OS Salmirol boliviensis boliviensis (Bolivian squirrel monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.  
 OX NCBI\_TaxID:39432;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Liu J.C., Makova K.D., Adkins R.M., Gibson S., Li W.H.;  
 RX MEDLINE:21265430; PubMed:11371582;  
 RT "Episodic evolution of growth hormone in primates and emergence of the  
 species specificity of human growth hormone receptor";  
 RL Mol. Biol. Evol. 18:945-953(2001).  
 CC -!- FUNCTION: Plays an important role in growth control. Its major  
 role in stimulating body growth is to stimulate the liver and  
 other tissues to secrete IGF-1. It stimulates both the  
 differentiation and proliferation of myoblasts. It also stimulates  
 amino acid uptake and protein synthesis in muscle and other  
 tissues (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL; AF339060; AAK62287.1;  
 DR InterPro; IPR001400; Somatotropin.  
 DR Pfam; PF00103; hormone; 1.  
 DR PRINTS; P00836; SOMATOTROPIN.  
 DR PROSITE; PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE; PS00338; SOMATOTROPIN\_2; 1.  
 KW Hormone; Pituitary; Signal.  
 FT SIGNAL 1 26 BY SIMILARITY.  
 FT CHAIN 27 217 SOMATOTROPIN.  
 FT DISULFID 79 191 BY SIMILARITY.  
 FT DISULFID 208 215 BY SIMILARITY.  
 SQ SEQUENCE 217 AA; 24864 MW; 951528992C529F7 CRC64;

Query Match 54.8%; Score 437; DB 1; Length 217;  
 Best Local Similarity 91.3%; Pred. No. 1e-35;  
 Matches 84; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 FPTPLSLFONAMLRHRLHQLAFDTYQEEAYIPKEQKYSFLONPQTSLSFSESPT 61  
 DB 27 FPTPLSLFONAMLRHRLHQLAFDTYQEEAYIPKEQKYSFLONPQTSLSFSESPT 86  
 QY 62 PSNRETOOKSNLELLRISLLLIQSWLEPVQ 93

DB 87 PASKKETOOKSNLELLRISLLLIQSWLEPVQ 118

RESULT 5  
 SOM2\_PANTR STANDARD: PRT: 217 AA.  
 ID P58757;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Growth hormone variant precursor (GH-V) (Placenta-specific growth  
 hormone) (Growth hormone 2).  
 GN GH2.  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 OX NCBI\_TaxID:9598;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Revol A., Esquivel D., Sanriago D., Bartera-Saldana H.;  
 RT "Independent duplication of the growth hormone gene in three  
 Anthropoid lineages";  
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Plays an important role in growth control. Its major  
 role in stimulating body growth is to stimulate the liver and  
 other tissues to secrete IGF-1. It stimulates both the  
 differentiation and proliferation of myoblasts. It also stimulates  
 amino acid uptake and protein synthesis in muscle and other  
 tissues.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Expressed in the placenta.  
 CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL; AF374233; AAL72285.1;  
 DR InterPro; IPR001400; Somatotropin.  
 DR Pfam; PF00103; hormone; 1.  
 DR PROSITE; PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE; PS00338; SOMATOTROPIN\_2; 1.  
 KW Hormone; Placenta; Signal; Glycoprotein.  
 FT SIGNAL 1 26 BY SIMILARITY.  
 FT CHAIN 27 217 GROWTH HORMONE VARIANT.  
 FT DISULFID 79 191 BY SIMILARITY.  
 FT DISULFID 208 215 BY SIMILARITY.  
 SQ SEQUENCE 217 AA; 24990 MW; 1592A429075677DE CRC64;

Query Match 54.5%; Score 434.5; DB 1; Length 217;  
 Best Local Similarity 78.9%; Pred. No. 1.8e-35;  
 Matches 90; Conservative 4; Mismatches 9; Indels 11; Gaps 1;

QY 2 FPTPLSLFONAMLRHRLHQLAFDTYQEEAYIPKEQKYSFLONPQTSLSFSESPT 61  
 DB 27 FPTPLSLFONAMLRHRLHQLAFDTYQEEAYIPKEQKYSFLONPQTSLSFSESPT 86  
 QY 62 PSNRETOOKSNLELLRISLLLIQSWLEPVQ 104  
 DB 87 PSNRVKTOKSNLELLRISLLLIQSWLEPVQ 140

RESULT 6  
 SOMA\_CALJA STANDARD: PRT: 217 AA.  
 ID SOMA\_CALJA  
 AC Q9GMB3;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)

**DT** 28-FEB-2003 (Rel. 41, Last annotation update)  
**RN** Somatotropin precursor (Growth hormone).  
GH1.  
**OS** Callithrix jacchus (Common marmoset).  
**OC** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
**CC** Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;  
**NCBI\_TaxID=9483;**

[1]  
**SEQUENCE FROM N.A.**  
**RA** Wallis O.C., Wallis M.;  
**RT** "Cloning and characterisation of a putative growth hormone encoding  
gene from the marmoset (*Callithrix jacchus*).";  
**RL** Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
**-!** FUNCTION: Plays an important role in growth control. Its major  
role in stimulating body growth is to stimulate the liver and  
other tissues to secrete IGF-1. It stimulates both the  
differentiation and proliferation of myoblasts. It also stimulates  
amino acid uptake and protein synthesis in muscle and other  
tissues (By similarity).

**-!** SUBCELLULAR LOCATION: Secreted.  
**-!** SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).

-----  
**EMBL:** AJ297563; CAC03481.1; .  
**HSSP:** P01241; 1A22.  
**InterPro:** IPRO01400; Somatotropin.  
**Pfam:** PF00103; hormone\_1; PROPIN\_1; 1.  
**PROSITE:** PS00266; SOMATOTROPIN\_1; 1.  
**PROSITE:** PS00338; SOMATOTROPIN\_2; 1.  
**Hormone:** Pituitary; Signal.  
**SIGNAL:** 1 26 BY SIMILARITY.  
**CHAIN:** 27 217 SOMATOTROPIN.  
**DISULFID:** 79 191 BY SIMILARITY.  
**DISULFID:** 208 215 BY SIMILARITY.  
**SEQUENCE:** 217 AA; 24959 MW; EI02151A12CP6192 CRC64;

Query Match 54.2%; Score 432; DB i: Length 217;  
Best Local Similarity 91.2%; Pred. No. 3.2e+35;  
Matches 83; Conservative 5; Mismatches 3; Indels 0; Gaps 0:  
  
QY 2 FPTPLSRFLNAMLRHLHOLAFTDYQEFEAYIPKEQKYSFLQNPTSLSIFSES IPT 61  
||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||  
Db 27 FPTPLSLRLLDNLMLRALRHQLAFDPDYQEFEAYIPKEQKYSFLQNPTSLSICFSIPT 86  
  
QY 62 PSNRRETOOKSNLEIRLSILLIQSWLPVQ 92  
| : : : : | ||||| | ||||| | ||||| | |||||  
Db 87 PASKRRETOOKSNLEIRLSILLIQSWFPVQ 117

RESULT 7  
SOM2\_HUMAN  
ID ID SOM2\_HUMAN STANDARD; PRT: 217 AA.  
AC P01242; P09587;  
DT 21-JUL-1986 (Rel. 01, Created)  
DI 28-FEB-2003 (Rel. 41, Last sequence update)  
DE DE Growth hormone variant precursor (GH-V) (Placenta-specific growth hormone) (Growth hormone 2).  
GN GH2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiinae; Homo.  
NCBI\_Taxid=9606;  
[1]  
**SEQUENCE FROM N.A.** (ISOFORM 1).

MEDLINE=83182010; PubMed=7169009;  
Seeburg P.H.:  
"The human growth hormone gene family: nucleotide sequences show recent divergence and predict a new polypeptide hormone.";  
DNA 1:239-249(1982).  
[2]  
SEQUENCE FROM N.A. (ISOFORM 1 AND 2).  
MEDLINE=88243769; PubMed=33057;  
Cooke N.E., Ray J.J., Emery J.J., Liebhauer S.A.:  
"Two distinct species of human growth hormone-related mRNA in the human placenta predict the expression of novel growth hormone proteins.";  
J. Biol. Chem. 263:9001-9006(1988).  
[3]  
SEQUENCE FROM N.A. (ISOFORM 1);  
MEDLINE=89024984; PubMed=2460050;  
Igout A., Scippo H.L., Frankeue F., Hennel G.:  
"Cloning and nucleotide sequence of placental hGH-v cDNA.";  
Arch. Int. Physiol. Biochim. 96:63-67(1988).  
[4]  
SEQUENCE FROM N.A.  
MEDLINE=89307277; PubMed=274760;  
Chen E.Y., Liao Y.C., Smith V.H., Barrera-Saldana H.A., Gelinas R.E., Seeburg P.H.:  
"The human growth hormone locus: nucleotide sequence, biology, and evolution.";  
Genomics 4:479-497(1989).  
[5]  
SEQUENCE FROM N.A.  
TISSUE=Placenta;  
MEDLINE=22388257; PubMed=12477932;  
Straussberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner K.H., Schaefer C.F., Bhat N.K., Altschul S.F., Zeeberg B.B., Jewett K.H., Moore T.M., Wang J., Siegfried D.P., Jordan H., Marudine K., Farmer A.A., Rubin G.M., Hong L., Diachenko L., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Tohiyuki S., Carninci P., Prange C.A., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Ketteman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Rodriquez R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.F., Schneider A., Schein J.E., Jones S.J.M., Maria M.A.:  
"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[6]  
REVIEW  
MEDLINE=93321812; PubMed=10393484;  
Baumann G.:  
"Growth hormone heterogeneity in human pituitary and plasma."  
Horm. Res. 51 Suppl. 1:2-6(1999).  
-!- FUNCTION: Plays an important role in growth control. Its major role in stimulating body growth is to stimulate the liver and other tissues to secrete IGF-1. It stimulates both the differentiation and proliferation of myoblasts. It also stimulates amino acid uptake and protein synthesis in muscle and other tissues.  
-!- SUBUNIT: Monomer, dimer, trimer, tetramer and pentamer, disulfide-linked or non-covalently associated, in homopolymeric and heteropolymetric combinations. Can also form a complex either with GHBP or with the alpha2-macroglobulin complex.  
-!- SUBCELLULAR LOCATION: Secreted.  
-!- ALTERNATIVE PRODUCTS:  
Event-Alternative splicing; Named isoforms-2:  
Name-1: Synonyms=GH-V1;  
IsoId=P01242-1; Sequence=Displayed;  
Name-2: Synonyms=GH-V2;



```

CC      isoId-P01242-2; Sequence-VSP_006203;
CC      Note-No experimental confirmation available;
CC      -1- TISSUE SPECIFICITY: Expressed in the placenta.
CC      -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: K00470; AAA98619.1;
CC      EMBL: J03756; AAB59547.1;
CC      EMBL: J03756; AAB59548.1;
CC      EMBL: M38451; AAA35891.1;
CC      EMBL: J03071; AAA2552.1;
CC      EMBL: BC020760; AAB20760.1;
CC      PIR: A28072; STH0V2.
CC      PIR: D32435; STH0V.
CC      HSSP: P01241; 1A22.
CC      Genew: HGNC:4262; GH2.
CC      MW: 139240;
CC      GO: GO:0005180; F:peptide hormone; TAS.
CC      InterPro: IPR001400; Somatotropin.
CC      Pfam: PF00103; hormone; 1.
CC      PROSITE: PS00266; SOMATOTROPIN_1; 1.
CC      PROSITE: PS00338; SOMATOTROPIN_2; 1.
CC      Hormone; Placenta; Signal: Glycoprotein; Alternative splicing;
CC      Polymorphism.
CC      SIGNAL      1      26
CC      CHAIN      27      217      GROWTH HORMONE VARIANT.
CC      DISULFID      79      191      BY SIMILARITY.
CC      DISULFID      208      215      N-LINKED (GLCNAC...) (POTENTIAL).
CC      CARBOHYD      166      166      RLEDSPTGQIFNOSYKFDTKSHNDAILKNYGLLYCFR
CC      VARSPLIC      153      217      KMDKYETFLRVQCSYVEGSGCF -> VRVAPGIPNPGAP
CC      LASRDWGEKHC: LFSSQALTOENSPSYFSLVNPVGLSLQ
CC      PGEGGKWNK: REQCPSPALLLFLHFAEAGRWQPPDWA
CC      DIOSVLOOV ( isoform 2).
CC      FTID-VSP_006203.
CC      R -> W (IN DBSNP:5389).
CC      FTID-VAR_014591.
CC      FT CONFLICT      109      109      I -> I (IN REF. 2).
CC      SEQUENCE      217 AA; 24999 MW; 7B9324698E822F96 Ck64.
CC      -----
CC      Query Match      53.5%; Score 426.5; DB 1; Length 217;
CC      Best Local Similarity 78.1%; Pred. No. 1.1e-34;
CC      Matches 89; Conservative 4; Mismatches 10; Indels 11; Gaps 1;
CC      -----
QY      2 FPTIPLSRFDNMLRAHRLHQLAFDTYOEFEAYIPKEQKYSFLQNPTSLSFSESIPT 61
DB      27 FPTIPLSRFDNMLRAHRLHQLAFDTYOEFEAYIPKEQKYSFLQNPTSLSFSESIPT 86
QY      62 PSNREETOOKSNLELLRISLLTIQSWLEPVLQGTGRFVNQHLGCSHLVEA-----LY 114
DB      87 PSNREETOOKSNLELLRISLLTIQSWLEPVLQGTGRFVNQHLGCSHLVEA-----LY 139
RESULT 8
SOM2_MACMU      STANDARD:      PRI: 217 AA.
ID      SOM2_MACMU
AC      Q07370; Q28494.
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DE      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Growth hormone variant precursor (GH-V) (Placenta-specific growth
DE      hormone) (Growth hormone 2).
DE      GH2.
OS      Macaca mulatta (Rhesus macaque).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;

```

```

OC      Cercopitheidae; Macaca.
OX      NCBI_TaxID-9544;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Golos T.G.;
RL      Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
PC      MEDLINE-94008724; PubMed-8404617;
RA      Golos T.G.; Durning M.; Fisher J.M.; Fowler P.D.;
RT      "Cloning of four growth hormone/chorionic somatomammotropin-related
RT      complementary deoxyribonucleic acids differentially expressed during
RT      pregnancy in the rhesus monkey placenta.";
RL      Endocrinology 133:1744-1752(1993).
CC      -1- FUNCTION: Plays an important role in growth control. Its major
CC      role in stimulating body growth is to stimulate the liver and
CC      other tissues to secrete IGF-1. It stimulates both the
CC      differentiation and proliferation of myoblasts. It also stimulates
CC      amino acid uptake and protein synthesis in muscle and other
CC      tissues.
CC      -1- SUBCELLULAR LOCATION: Secreted (By similarity).
CC      -1- TISSUE SPECIFICITY: Expressed in the placenta.
CC      -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: U02293; AAA03391.1;
CC      EMBL: L16555; AAA20180.1;
CC      PIR: I67411; I67411.
CC      HSSP: P01241; 1HG0.
CC      InterPro: IPR001400; Somatotropin.
CC      Pfam: PF00103; hormone; 1.
CC      PRINTS: PR00836; SOMATOTROPIN.
CC      PROSITE: PS00266; SOMATOTROPIN_1; 1.
CC      PROSITE: PS00338; SOMATOTROPIN_2; 1.
CC      Hormone; Placenta; Signal: Glycoprotein.
CC      SIGNAL      1      26      BY SIMILARITY.
CC      CHAIN      27      217      GROWTH HORMONE VARIANT.
CC      DISULFID      79      191      BY SIMILARITY.
CC      DISULFID      208      215      BY SIMILARITY.
CC      CONFLICT      57      57      L -> F (IN REF. 2).
CC      CONFLICT      152      152      E -> G (IN REF. 2).
CC      SEQUENCE      217 AA; 25221 MW; 8DB116CBC24EA090 Ck64;
CC      -----
CC      Query Match      50.1%; Score 399; DB 1; Length 217;
CC      Best Local Similarity 66.9%; Pred. No. 5.4e-32;
CC      Matches 85; Conservative 9; Mismatches 19; Indels 14; Gaps 2;
CC      -----
QY      2 FPTIPLSRFDNMLRAHRLHQLAFDTYOEFEAYIPKEQKYSFLQNPTSLSFSESIPT 61
DB      27 FPTIPLSRFDNMLRAHRLHQLAFDTYOEFEAYIPKEQKYSFLQNPTSLSFSESIPT 86
QY      62 PSNREETOOKSNLELLRISLLTIQSWLEPVLQGTGRFVNQHLGCSHLVEA-----LY 114
DB      87 PSNREETOOKSNLELLRISLLTIQSWLEPVLQGTGRFVNQHLGCSHLVEA-----LY 139
QY      115 LVCGERG 121
DB      140 LKKLEEG 146
RESULT 9
PLL_HUMAN
ID      PLL_HUMAN
AC      P01243;
DT      21-JUL-1986 (Rel. 01, Created)

```



27 FPAMPSSLSFANAVLAHQHLLQALADTTKFERATYIPGGORYS-IGNAQATYCFSETIPA 81  
 62 PSNREETQOKSNLELLRISLLIQSWLEPVQ 92  
 86 PTGKEAQORSUMELLRFSLLLIQSWLGPVQ 116

RESULT 11  
 SOMA\_BALBO  
 ID SOMA\_BALBO STANDARD; PRT: 190 AA.  
 AC P33092;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Somatotropin (Growth hormone).  
 GN GHI.  
 OS Balaenoptera borealis (Sei whale).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;  
 OC Balaenopteridae; Balaenoptera.  
 ON NCBI\_TaxID-9768;  
 RN 11  
 RP SEQUENCE  
 RX MEDLINE-83000569; PubMed-7115813;  
 RY Yudaev N.A., Pankov Y.A., Hulatov A.A., Osipova T.A.:  
 RT "Amino acid sequence of seiwhale somatotropin.";  
 RL Biochimica 47:1059-1069(1982).  
 RN [2]  
 RP PRELIMINARY PARTIAL SEQUENCE.  
 RX Osipova T.A., Hulatov A.A., Pankov Y.A.:  
 RT "Structural studies of tryptic peptides from large cyanogen bromide  
 RL fragments of sei whale (Balaenoptera borealis) somatotropin.";  
 RL Biorg. Khim. 4:1589-1599(1978).  
 CC !- FUNCTION: Plays an important role in growth control. Its major  
 CC role in stimulating body growth is to stimulate the liver and  
 CC other tissues to secrete IGF-1. It stimulates both the  
 CC differentiation and proliferation of myoblasts. It also stimulates  
 CC amino acid uptake and protein synthesis in muscle and other  
 CC tissues.  
 CC !- SUBCELLULAR LOCATION: Secreted.  
 CC !- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
 DR PIR: P01040; P01040.  
 DR HSP: P01241; IAXI.  
 DR InterPro: IPR0C1400; Somatotropin.  
 DR Pfam: PF00103; Hormone\_1.  
 DR PROSITE: PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE: PS00338; SOMATOTROPIN\_2; 1.  
 KW Hormone; Pituitary.  
 FT DISULFID 52 163 HY SIMILARITY.  
 FT DISULFID 180 188 HY SEMIFUNCTIONALITY.  
 SQ SEQUENCE 190 AA; 21835 MW; 09F6FF6D814A75D6 CRC64;  
 Query Match 38.6%; Score 307.5; DB 1; Length 190;  
 Best Local Similarity 67.0%; Pred. No. 4.3e-23;  
 Matches 61; Conservative 14; Mismatches 15; Indels 1; Gaps 1

2 FPTPLSLRLDNLRAHLRLHAFDTYQEFEEAYIPKEQKYSFLONPOTLSLFSSEIPT 61  
 1 FPAMPSSLSFANAVLAHQHLLQALADTTKFERATYIPGGORY-FLONAQSTGCFSEVPT 59

62 PSNREETQOKSNLELLRISLLIQSWLEPVQ 92  
 60 PANDEAQORSVDVLELLRFSLLLIQSWLGPVQ 90

RESULT 12  
 SOMA\_HORSE  
 ID SOMA\_HORSE STANDARD; PRT: 216 AA.  
 AC F01245;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

QY	2	FPTILSLRFDNAMLRAHRLHQLAFDPTQYQEFEEAYIPKEQKYSFLQNPQTSLSFSSEIPT 61
DB	27	FPAMPFLSSILFANAVLRAQHLHQLAADTYKEFERAYIPEGQYRS-IQNAQAACFSETIPA 85
QY	62	PSNREETOOKSNLELLRLISLILIOSWLEPVOL 93
DB	86	PTGKDEAQRSDMELLRLISLILIOSWLEPVOL 117
<p>RESULT 13</p> <p>SOMA_GALSE STANDARD: PRT. 217 AA.</p> <p>AC Q9GKAL;</p> <p>DT 28-FEB-2003 (Rel. 41, Created)</p> <p>DT 28-FEB-2003 (Rel. 41, Last sequence update)</p> <p>DT 28-FEB-2003 (Rel. 41, Last annotation update)</p> <p>DE Somatotropin precursor (growth hormone)</p> <p>GN GH1.</p> <p>OS Galago senegalensis (Northern lesser bushbaby).</p> <p>OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;</p> <p>OC Mammalia; Eutheria; Primates; Strepsirhini; Galagonidae; Galago.</p> <p>OX NCBI_TaxID=9465;</p> <p>RN [1]</p> <p>RP SEQUENCE FROM N.A.</p> <p>RX PubMed=11141192;</p> <p>RA Adkins R.M., Nekrutenko A., Li W.-H.;</p> <p>RT "Bushbaby growth hormone is much more similar to nonprimate growth hormones than to rhesus monkey and human growth hormones."</p> <p>RL Mol. Biol. Evol. 18:55-61(2001).</p> <p>CC !- FUNCTION: Plays an important role in growth control. Its major role in stimulating body growth is to stimulate the liver and other tissues to secrete IGF-1. It stimulates both the differentiation and proliferation of myoblasts. It also stimulates amino acid uptake and protein synthesis in muscle and other tissues.</p> <p>CC !- SUBCELLULAR LOCATION: Secreted.</p> <p>CC !- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.</p> <p>CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a>)</p> <p>EMBL: AF292938; AAC44952.1;</p> <p>DR HSSP: P01246; LBST.</p> <p>DR InterPro: IPR001400; Somatotropin.</p> <p>DR Pfam: PF00103; hormone.</p> <p>DR PRINTS: PR00836; SOMATOTROPIN.</p> <p>DR PROSITE: PS00266; SOMATOTROPIN_1; 1.</p> <p>DR PROSITE: PS00338; SOMATOTROPIN_2; 1.</p> <p>KW Hormone; Pituitary; Signal.</p> <p>FT SIGNAL 1 26 BY SIMILARITY.</p> <p>FT CHAIN 27 217 SOMATOTROPIN.</p> <p>FT DISULFID 79 190 BY SIMILARITY.</p> <p>FT DISULFID 207 215 BY SIMILARITY.</p> <p>SQ SEQUENCE 217 AA; 24481 MW; 2FB61C031136F005 CRC64;</p> <p>Query Match 38.5%; Score 306.5; DB 1; Length 217;</p> <p>Best Local Similarity 65.2%; Pred. No. 6.2e-23;</p> <p>Matches 60; Conservative 14; Mismatches 17; Indels 1; Gaps 1;</p>		
QY	2	FPTILSLRFDNAMLRAHRLHQLAFDPTQYQEFEEAYIPKEQKYSFLQNPQTSLSFSSEIPT 61
DB	28	FPAMPFLSSILFANAVLRAQHLHQLAADTYKEFERAYIPEGQYRS-IQNAQAACFSETIPA 86
QY	62	PSNREETOOKSNLELLRLISLILIOSWLEPVOL 93
DB	87	PTGKDEAQRSDMELLRLISLILIOSWLEPVOL 118

## RESULT 14

SOMA\_NYCPY  
ID SOMA\_NYCPY STANDARD; PRT; 217 AA.  
AC Q9GMB2;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Somatotropin precursor (Growth hormone).  
GN GH1.  
OS Nycticebus pygmaeus (Pygmy slow loris).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Strepsirhini; Loridae; Nycticebus.  
OX NCBI\_TaxID=101278;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Wallis O.C., Zhang Y.P., Wallis M.;  
RT "Cloning and characterisation of the gene encoding slow loris growth hormone."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
CC 1- FUNCTION: Plays an important role in growth control. Its major role in stimulating body growth is to stimulate the liver and other tissues to secrete IGF-1. It stimulates both the differentiation and proliferation of myoblasts. It also stimulates amino acid uptake and protein synthesis in muscle and other tissues.  
CC 1- SUBCELLULAR LOCATION: Secreted.  
CC 1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; AJ297562; CAC03504.1;  
CC HSSP; P01246; 1BST.  
CC InterPro; IPR001400; Somatotropin.  
CC Pfam; PF00103; hormone; 1.  
CC PRINTS; PR00836; SOMATOTROPIN.  
CC PROSITE; PS00266; SOMATOTROPIN\_1; 1.  
CC PROSITE; PS00338; SOMATOTROPIN\_2; 1.  
KW Hormone; Pituitary; Signal.  
FT SIGNAL 1 27 BY SIMILARITY.  
FT CHAIN 28 217 SOMATOTROPIN.  
FT DISULFID 79 190 BY SIMILARITY.  
FT DISULFID 207 215 BY SIMILARITY.  
SQ SEQUENCE 217 AA: 24395 MW: 7FE90D7FE59085F6 CRC64;

Query Match 38.5%; Score 306.5; DB 1; Length 217;  
Best Local Similarity 65.2%; Pred. No. 6.2e-23;  
Matches 60; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

QY 2 FPTPLSLRFDNMLRAHRLHQLADTYQEFEEAYIPKEQKYSFLQNPTSLSFSESIPT 61  
DB 28 FPAMPFLSLFANVLAHQHLHQLADTYKEFERAYIPKEQKYS-IGNAQAAPCFSETIPA 86  
QY 62 PSNREETOOKSNLELLRLSLLLIQSWLEPVQ 93  
DB 87 PTGKEAQOQSDMELLRFSLLLIQLSWLGPVL 118

## RESULT 15

SOMA\_MOUSE  
ID SOMA\_MOUSE STANDARD; PRT; 216 AA.  
AC P06980;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Somatotropin precursor (Growth hormone).  
GN GH1 OR GH.

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85261358; PubMed=2991252;  
RA Linzer D.I.H., Talamantes F.;  
RT "Nucleotide sequence of mouse prolactin and growth hormone mRNAs and expression of these mRNAs during pregnancy."  
RL J. Biol. Chem. 260:9574-9579(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FZTDU; TISSUE=Liver;  
RX MEDLINE=96194803; PubMed=86 48;  
RA Das P., Meyer L., Seyfert H., Brockmann G., Schwerin M.;  
RT "Structure of the growth hormone-encoding gene and its promoter in mice."  
RL Gene 169:209-213(1996).  
CC 1- FUNCTION: Plays an important role in growth control. Its major role in stimulating body growth is to stimulate the liver and other tissues to secrete IGF-1. It stimulates both the differentiation and proliferation of myoblasts. It also stimulates amino acid uptake and protein synthesis in muscle and other tissues.  
CC 1- SUBCELLULAR LOCATION: Secreted.  
CC 1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X02891; CAA26650.1;  
CC EMBL; Z46663; CAA86658.1;  
CC PIR; B23911; STMS.  
CC HSSP; P01246; 1BST.  
CC MGD; MGI:95707; Gh.  
CC InterPro; IPR001400; Somatotropin.  
CC Pfam; PF00103; hormone; 1.  
CC PROSITE; PS00266; SOMATOTROPIN\_1; 1.  
CC PROSITE; PS00338; SOMATOTROPIN\_2; 1.  
KW Hormone; Pituitary; Signal.  
FT SIGNAL 1 26 BY SIMILARITY.  
FT CHAIN 27 216 SOMATOTROPIN.  
FT DISULFID 78 184 BY SIMILARITY.  
FT DISULFID 206 214 BY SIMILARITY.  
SQ SEQUENCE 216 AA: 24716 MW: 98666A1AE25D65FC CRC64;

Query Match 38.2%; Score 304.5; DB 1; Length 216;  
Best Local Similarity 64.8%; Pred. No. 9.7e-23;  
Matches 59; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

QY 2 FPTPLSLRFDNMLRAHRLHQLADTYQEFEEAYIPKEQKYSFLQNPTSLSFSESIPT 61  
DB 27 FPAMPFLSLFANVLAHQHLHQLADTYKEFERAYIPKEQKYS-IGNAQAAPCFSETIPA 85  
QY 62 PSNREETOOKSNLELLRLSLLLIQSWLEPVQ 92  
DB 86 PTGKEAQOQSDMELLRFSLLLIQLSWLGPVQ 116

Search completed: September 16, 2003, 12:38:52  
Job time : 12.6732 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: September 16, 2003, 12:34:00 : Search time 36.1868 Seconds  
(without alignments)  
1069.670 Million cell updates/sec

Title: US-09-423-100-7  
Perfect score: 797  
Sequence: 1 MEPTPLSRFDNMLRAHR.....IVEQCCTICSLYOLENYCN 150

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues  
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL23.\*
- 1: sp\_archea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phage.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*
  - 15: sp\_virus.\*
  - 16: sp\_bacteriaph.\*
  - 17: sp\_archeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	436	54.7	217	6	Q8WNE0	Q8WNE0 ateles geof
2	427.5	53.6	245	4	O14644	O14644 homo sapien
3	407.5	51.1	217	6	Q07369	Q07369 macaca mula
4	396	49.7	212	6	Q07368	Q07368 macaca mula
5	396	49.7	217	6	Q07367	Q07367 macaca mula
6	381	47.8	217	4	O14407	O14407 homo sapien
7	348	43.7	217	6	Q8WNE9	Q8WNE9 ateles geof
8	341	42.8	202	4	O14643	O14643 homo sapien
9	322.5	40.5	217	6	Q8M174	Q8M174 callithrix
10	306.5	38.5	216	11	O70615	O70615 spalax leuc
11	301.5	37.8	216	6	Q8M173	Q8M173 delphinus d
12	301.5	37.8	216	6	Q8HYE5	Q8HYE5 ailuropoda
13	298.5	37.5	216	11	Q9R2C3	Q9R2C3 mus musculu
14	297.5	37.3	204	6	Q95205	Q95205 ovis aries
15	297.5	37.3	216	11	Q9JKM4	Q9JKM4 cavia porce
16	297	37.3	217	6	Q8M175	Q8M175 callithrix

17	290.5	36.4	192	6	Q9TU21	Q9TU21 capra hircu
18	289.5	36.3	192	6	Q9TQW9	Q9TQW9 bos indicu
19	287.5	36.1	190	11	Q9JKG0	Q9JKG0 cavia porce
20	286.5	35.9	178	6	Q95MJ5	Q95MJ5 tarsius ban
21	285.5	35.8	217	6	Q9BEC0	Q9BEC0 tarsius ja
22	285.5	35.8	217	6	Q9BEB9	Q9BEB9 tarsius ja
23	285	35.8	167	4	78451	78451 homo sapien
24	283.5	35.6	178	6	Q95MJ6	Q95MJ6 tarsius syr
25	280.5	35.2	217	6	Q28957	Q28957 sus scrofa
26	271.5	34.1	110	6	Q8HXV2	Q8HXV2 pongo pygma
27	266.5	33.4	110	11	Q91X13	Q91X13 spermophilu
28	265.5	33.3	143	6	Q95240	Q95240 canis fami
29	261.5	32.8	218	13	Q9PU72	Q9PU72 cynops pyrr
30	254	31.9	110	6	Q8WNE6	Q8WNE6 felis silve
31	246.5	30.9	145	6	Q9BDR4	Q9BDR4 galago cras
32	234	29.4	199	4	O14406	O14406 homo sapien
33	233.5	29.3	195	13	Q91386	Q91386 amia calva
34	224.5	28.2	106	13	Q91807	Q91807 rana pipien
35	201.5	25.3	110	13	Q98TA8	Q98TA8 pantodon bu
36	197	24.7	108	13	Q9DDE5	Q9DDE5 brachydantio
37	195.5	24.5	108	13	Q907N4	Q907N4 catla catla
38	195	24.5	111	13	Q98TB0	Q98TB0 chitala chi
39	193.5	24.3	110	13	Q902Y1	Q902Y1 hiodon alos
40	191.5	24.0	111	13	Q98TA7	Q98TA7 osteoqllossu
41	187.5	23.5	87	13	Q98TA9	Q98TA9 gnathonemus
42	187.5	23.5	108	13	Q98TB1	Q98TB1 catostomus
43	186.5	23.4	93	6	Q8HXV8	Q8HXV8 bos mutus g
44	185.5	23.3	91	13	Q98TB2	Q98TB2 ambloplites
45	177.5	22.3	209	13	Q8AXX9	Q8AXX9 anquilla an

ALIGNMENTS

RESULT 1  
Q8WNE0  
ID Q8WNE0 PRELIMINARY, PRT: 217 AA.  
AC Q8WNE0  
DT 01-MAR-2002 (TRENBLrel. 20. Created)  
DT 01-MAR-2002 (TRENBLrel. 20. Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23. Last annotation update)  
DE Growth hormone.  
GN GH-N.  
OS Ateles geoffroyi (Black-handed spider monkey).  
OC Eukaryota; Metazoa; Chordata; Cladocera; Vertebrata; Euteleostomi;  
OF Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Ateles.  
OX NCBI TaxID:9559;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Revol A., Esquivel D., Santiago D., Barrera-Saldana H.  
RT \*Independent duplication of the growth hormone gene in three  
RT Anthropoid lineages.\*  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF374234; AAL72286.1; -  
DR InterPro; IPR001400; Somatotropin.  
DR Pfam; PF00103; hormone; 1.  
DR PROSITE; PS00266; SOMATOTROPIN\_1; 1.  
DR PROSITE; PS00338; SOMATOTROPIN\_2; 1.  
DR PROSITE; PS00338; SOMATOTROPIN\_2; 1.  
SQ SEQUENCE 217 AA: 24894 MW: 425829FF41EEAE6 CRC64;

Query Match 54.7%; Score 436; DB 6; Length 217;  
Best Local Similarity 66.9%; Pred. No. 1.4e-38;  
Matches 97; Conservative 8; Mismatches 24; Indels 16; Gaps 4;  
QY 2 FPTPLSRFDNMLRAHRHLQALFDYVQEEFAYIPKEQKYSFLONPQTSLSFSESPT 61  
Db 27 FPTPLSRFDNMLRAHRHLQALFDYVQEEFAYIPKEQKYSFLONPQTSLSFSESPT 86  
QY 62 PSNREETOQKSNLELLRLISLLIQSWLEPVLGTGPRFVNOHLCG-----SHLVE 111  
Db 87 PASKKETQKSNLELLRLISLLIQSWLEPVLGTGPRFVNOHLCG-----SHLVE 145  
QY 112 ALYLVCG--ERGFYTPKTRGIVEQ 134

```

Db 146 GIOTLMORLEDG---SPUTGEIFRQ 167

RESULT 2
O14644
AC O14644 PRELIMINARY; PRT: 245 AA.
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Placental growth hormone isoform hGH-V3 precursor.
GN HGH-V.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Term placenta;
RA MEDLINE=94008724; PubMed=9709963;
RA Boguszewski C.L., Svensson P.A., Jansson T., Clark R.,
RA Carlsson L.M.S., Carlsson B.;
RT *Cloning of two novel growth hormone transcripts expressed in human
RT placenta.*;
RL J. Clin. Endocrinol. Metab. 83:2878-2885(1998).
RL EMBL: AF006061; AAB71829.1;
DR HSSP: P01241; IAZ2.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
KW Signal.
FT SIGNAL 1 26 POTENTIAL
SQ SEQUENCE 245 AA: 27101 MW: 14CC7F8CD75D91C8 CRC64:

Query Match 53.6%; Score 427.5; DB 4; Length 245;
Best Local Similarity 71.9%; Pred. No. 1.3e-37;
Matches 92; Conservative 6; Mismatches 17; Indels 13; Gaps 2;

OY 2 FPTPLSLFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSFSESIPTP 61
|||||
DB 27 FPTPLSLFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSFSESIPTP 86
|||||
OY 62 PSNREETOQKSNLELLRISLLIQLSWLEPVQ-----GTGPRFVNOHLGSHIVL 110
|||||
DB 87 PSNRVKTQKSNLELLRISLLIQLSWLEPVQLKSVFANSLVVGASDSNVYRHL---KDLF 144
|||||
OY 111 EALYLVCG 118
|||||
DB 145 EGIOTLLIC 152
|||||

RESULT 3
Q07369 PRELIMINARY; PRT: 217 AA.
AC Q07369;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Chorionic somatomammotropin-3.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Midpregnancy placenta;
RA MEDLINE=94008724; PubMed=8404617;
RA Golos T.G., Durning M., Fisher J.M., Fowler P.D.;
RT "Cloning of four growth hormone/chorionic somatomammotropin-related
RT complementary deoxyribonucleic acids differentially expressed during
RT pregnancy in the rhesus monkey placenta.";
RL Endocrinology 133:1744-1752(1993).
DR EMBL: L16553; AAA18840.1;
DR HSSP: P01241; IAXI.
DR InterPro: IPR00400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PROSITE: PS00338; SOMATOTROPIN_1; 1.
FT NON_TER 1
SQ SEQUENCE 212 AA: 24526 MW: 27809110K256EAF5 CRC64:

Query Match 49.7%; Score 396; DB 6; Length 212;
Best Local Similarity 82.2%; Pred. No. 2.6e-34;
Matches 74; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

OY 3 PTIPLSLFDNAMLRAHRLHQLAFDTYQFEFEAYIPKEQKYSFLONPOTSLSFSESIPTP 62
|||||
DB 23 PVPPLSLRFLDHAMIQAHRLHQLAFDTYQFEFEAYIPKEKKHSLMKNPQASFCFADSIPTP 82
|||||
OY 13 SNREETOQKSNLELLRISLLIQLSWLEPVQ 92
|||||
DB 83 SNLEETOQKSNLELLRISI LLIQSWLEPVQ 112
|||||

RESULT 5
Q07367 PRELIMINARY; PRT: 217 AA.
AC Q07367;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Chorionic somatomammotropin-1.

```

```

OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Midpregnancy placenta.
RX MEDLINE=94008724; PubMed=8404617.
RA Golos T.G., Durning M., Fisher J.M., Fowler P.D.:
RT "Cloning of four growth hormone/chorionic somatomammotropin-related
RT complementary deoxyribonucleic acids differentially expressed during
RT pregnancy in the rhesus monkey placenta.";
RL Endocrinology 133:1744-1752(1993).
DR EMBL: L16552; AAA18839.1; -.
DR HSSP: P01241; JAXI
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
SQ SEQUENCE 217 AA: 24942 MW: 55289.15 131F2BC CRC64:

Query Match 49.7%; Score 396; DB 6; Length 217;
Best Local Similarity 82.2%; Pred. No. 2.7e-34;
Matches 74; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

QY 3 PTIPLSRFLDNAMLRHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPTSLSESIPTP 62
DL 28 PVPPLSRFLDNAMLRHRLHQLAFDTYQEFEEAYIPKEKHSLSLWENPQASFCFADSIPTP 87
QY 63 SNREETOQKSNLELLRLSLLLIQSWLEPVQ 92
DB 88 SNLEETOQKSNLELLRLSLLLIQSWLEPVQ 117

RESULT 6
Q14407
ID Q14407 PRELIMINARY; PRT: 217 AA.
AC Q14407;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Chorionic somatomammotropin CS-2 (Chorionic somatomammotropin hormone
DE 2).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89307277; PubMed=2744760;
RA Chen E.Y., Liao Y.C., Smith D.H., Barrera-Saldana H.A., Gelin R.E.,
RA Seeburg P.H.:
RT "The human growth hormone locus: nucleotide sequence, biology, and
RT evolution.";
RL Genomics 4:479-497(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91102558; PubMed=1980158;
RA Vnencak-Jones C.L., Phillips J.A., III.:
RT "Hot spots for growth hormone gene deletions in homologous regions
RT outside of Alu repeats.";
RL Science 250:1745-1748(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.:
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.:
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

```

```

DR EMBL: J03071; AAA52553.1; -.
DR EMBL: BC022044; AAH22044.1; -.
DR EMBL: BC035965; AAH35965.1; -.
DR HSSP: P01241; IA22.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
SQ SEQUENCE 217 AA: 24994 MW: 39540.00 86B2E95; CRC64:

Query Match 47.8%; Score 381; DB 4; Length 217;
Best Local Similarity 82.0%; Pred. No. 1.1e-32;
Matches 73; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 4 TIPLSKFLDNAMLRHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPTSLSESIPTP 63
DB 29 TVPLSKFLDNAMLRHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPTSLSESIPTP 88
QY 64 NREETOQKSNLELLRLSLLLIQSWLEPVQ 92
DB 89 NREETOQKSNLELLRLSLLLIQSWLEPVQ 117

RESULT 7
Q8WND9
ID Q8WND9 PRELIMINARY; PRT: 217 AA.
AC Q8WND9;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Growth hormone.
GN GH-V.
OS Ateles geoffroyi (Black-handed spider monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Ateles.
OX NCBI_TaxID=9509;
RN [1]
RP SEQUENCE FROM N.A.
RA Revol A., Esquivel B., Santiago B., Barrera-Saldana H.:
RT "Independent duplication of the growth hormone gene in three
RT Anthropoid lineages.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF374235; AAL72287.1; -.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
SQ SEQUENCE 217 AA: 25293 MW: 74174.51 875053E CRC64:

Query Match 43.7%; Score 348; DB 6; Length 217;
Best Local Similarity 75.8%; Pred. No. 3.7e-29;
Matches 69; Conservative 8; Mismatches 14; Indels 0; Gaps 0;

QY 2 FTIPLSRFLDNAMLRHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPTSLSESIPT 61
DB 27 FPIPLSRFLFGDMLRAHRLHQLAFDTYQELLENCIPKKQKFFLNRKPNKFLCFSESIPT 86
QY 62 PSNREETOQKSNLELLRLSLLLIQSWLEPVQ 92
DB 87 PFNKEVLAKSLELLHLS LLIQSWLEPVQ 117

RESULT 8
O14643
ID O14643 PRELIMINARY; PRT: 202 AA.
AC O14643;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Placental growth hormone 20kDa isoform precursor.
GN HGH-V

```



QY	63	SNRETOQKSNLELLRLIRLLLIQSNLEPVOLGTGPRFVQHL	104
DB	88	PHKEEMUGKSNVELLHISLLLIQSNLEPMO-RLGSIANSOL	128
RESULT 10			
QY	070615	PRELIMINARY;	PRT: 216 AA.
ID	070615	PRELIMINARY;	PRT: 216 AA.
AC	070615;		
DT	01-AUG-1998 (TREMBLrel. 07, Created)		
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)		
DT	01-MAR-2003 (TREMBLrel. 27, Last annotation update)		
DE	Growth hormone precursor		
OS	Spalax leucodon ehrenbergi (Ehrenberg's mole rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Spalacinae;		
OC	Nannospalax.		
OX	NCBI_TaxID=30637;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99124645; PubMed=9426177;		
RA	Lioupis A., Nevo E., Wallis M.,		
RT	"Cloning and characterisation of the gene encoding mole rat (Spalax		
RT	ehrenbergi) growth hormone."		
RL	J. Mol. Endocrinol. 22:29-36(1999).		
DR	EMBL: AJ005819; CAA06716.1;		
DR	HSSP: P01241; IAXI.		
DR	InterPro: IPR001400; Somatot pin.		
DR	Pfam: PF00103; hormone; 1.		
DR	PRINTS: PR00836; SOMATOTROPIN		
DR	PROSITE: PS00266; SOMATOTROPIN_1;		
DR	PROSITE: PS00338; SOMATOTROPIN_2;		
KW	Signal.		
FT	SIGNAL 1 26		POTENTIAL.
FT	CHAIN 27 216		GROWTH HORMONE.
SQ	SEQUENCE 216 AA; 24627 MW; REAB8A523BA0ADFE CRC64;		
Query Match 38.5%; Score 306.5; DB 11; Length 216;			
Best local Similarity 65.9%; Pred. No. 1e-24;			
Matches 60; Conservative 13; Mismatches 17; Indels 1; Gaps			
QY	2	FTPTLSLFLONAMIKAKRLQYAFDTYQFEFEAYIPKEQYSLQNPQTSLSFSRSTIP	
DB	27	FPAPMLSNLFANAVLRAGHLHQIAADTYKEFERAYIPQGRYS-IQNAQAACFSETIP	
QY	62	FSNRRETOQKSNLELLRLIRLLLIQSNLEPVQ	92
DB	86	PTKRETAQGRSINLEIRISLLIQSNLEPVQ	116
RESULT 11			
QY	08M173	PRELIMINARY;	PRT: 216 AA.
ID	08M173	PRELIMINARY;	PRT: 216 AA.
AC	08M173;		
DT	01-OCT-2002 (TREMBLrel. 22, Created)		
DT	01-OCT-2002 (TREMBLrel. 22, Last sequence update)		
DT	01-MAR-2003 (TREMBLrel. 27, Last annotation update)		
DE	Growth hormone precursor.		
OS	GN.		
OC	Delphinus delphis (Saddleback dolphin) (Black sea dolphin).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Delphinida		
OX	NCBI_TaxID=9728;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Liver;		
RA	Manion Z., Wallis O.C., Wallis M.;		
RT	"Cloning and characterisation of the GH gene from the common dolphin		
RT	(Delphinus delphis)."		
RL	Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AJ492191; CAD37292.1;		

RESULT 15  
Q9JKM4

```

ID Q9JKM4 PRELIMINARY; PRT: 216 AA.
AC Q9JKM4;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Growth hormone precursor.
OS Cavia porcellus (Guinea pig);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RN SEQUENCE FROM N.A.
RA Odjirico D.M., Fuller P.J., Herington A.C.;
RT "Cloning and sequence of guinea pig growth hormone (GH).";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF233853; AAF36409.1;
DR HSSP: P01241; LAXI.
DR InterPro: IPR001400; Somatotropin.
DR Pfam: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
KW Signal.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 216 GROWTH HORMONE
SQ SEQUENCE 216 AA; 24822 MW; 459968BELL9B08DD3 CRC64;

Query Match 37.3%; Score 297.5; DB 11; Length 216;
Best Local Similarity 63.7%; Pred. No. 9.2e-24;
Matches 58; Conservative 14; Mismatches 18; Indels 1; Gaps 1;

QY 2 FPTPLSLFLDNLRAHLHQLAFDTYCEFEAYIPKEQKYSFLONPQTSLSFSESIPT 61
Db 27 FRAMPSSSLFGNAVLRQAHLQAADTYKEFRTYIPECQKYS-IHNTQTAFCESETIPA 85

QY 62 PSNRETOOKSNLELLRISLLLIQSWLEPVQ 92
Db 86 PTDKEAQQRSDVELLHFSLLLIQSWLGPVQ 116

```

Search completed: September 16, 2003, 12:40:01  
Job time : 37.1868 secs